The implicit function theorem revisited

by M. Shub

This paper is a survey of some results on Newton's method as applied to the implicit function theorem, homotopy methods, and Bezout's theorem. An application to macroeconomics is also described.

Introduction

Newton's method is one of the primary techniques used for solving systems of nonlinear equations. It attempts to replace the problem of solving nonlinear equations by an iterative process of linear-equation solving. First, we introduce some terminology. If we have n equations in m real variables $[f_1(x_1, \cdots, x_m) = 0, \cdots, f_n(x_1, \cdots, x_m) = 0]$, we can summarize this information by one vector equation,

$$\mathbf{f}(\mathbf{x}) = \mathbf{0},$$

where $\mathbf{x} \equiv (x_1, \cdots, x_m)$ and $\mathbf{f} \equiv (f_1, \cdots, f_n)$. Thus, $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$. The symbol $\mathbb{R}^m$ stands for m-dimensional Euclidean space; the norm of a vector in this space is $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_m^2}$. Sometimes our problems have constraints, as is apparent in our examples from economics, below. So instead of considering only $\mathbb{R}^m$ or $\mathbb{R}^n$, we consider more generally normed linear spaces (over the reals or complexes), which we call $\mathbf{E}$ and $\mathbf{F}$, respectively. (A linear space has to do with vectors, and a norm measures their lengths.) Also, sometimes our functions are not globally defined; that is, they are not defined on all of $\mathbf{E}$, but only in some local region. This may occur, for example, because of the failure of a power series to converge or because a denominator becomes zero.

The linear system by which the nonlinear equations are replaced in Newton's method is given by the derivatives of $\mathbf{f}$ at a point $\mathbf{x}$, which we denote $D\mathbf{f}(\mathbf{x})$. For $\mathbb{R}^m$ and $\mathbb{R}^n$ as above, $D\mathbf{f}(\mathbf{x})$ is the Jacobian matrix

$$\left(\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right) \qquad i = 1, \cdots, n, \qquad j = 1, \cdots, m.$$

In summary, Newton's method is employed to solve nonlinear equations $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, where $\mathbf{f} : \mathbf{E} \to \mathbf{F}$ is a differentiable function between two normed linear spaces (see **Figure 1**), and $\mathbf{f}$ is either globally or locally defined.

For an initial point $\mathbf{x}_0$ in $\mathbf{E}$, Newton's method solves the linearized equation $D\mathbf{f}(\mathbf{x}_0)\mathbf{v} = -\mathbf{f}(\mathbf{x}_0)$ for $\mathbf{v}$ and replaces $\mathbf{x}_0$ with $\mathbf{x}_0 + \mathbf{v}$ as the initial point for the next iteration. Iterations are terminated when some error criterion is satisfied. Usually the derivative $D\mathbf{f}(\mathbf{x}_0)$ is assumed to be invertible; then we can write Newton's method as

$$\mathbf{x} \to \mathbf{N}_\mathbf{f}(\mathbf{x}), \text{ where } \mathbf{N}_\mathbf{f}(\mathbf{x}) \equiv \mathbf{x} - (D\mathbf{f}(\mathbf{x}))^{-1}\mathbf{f}(\mathbf{x}). \tag{1}$$

Thus, $\mathbf{N}_\mathbf{f}(\mathbf{x})$ is a transformation from the space $\mathbf{E}$ to itself. Both $\mathbf{E}$ and $\mathbf{F}$ may be Euclidean m-space, $\mathbb{R}^m$. When $\mathbf{E}$ is the real numbers, $\mathbb{R}$, then (1) has the familiar form

$$\mathbf{x} \to \mathbf{x} - \frac{\mathbf{f}(\mathbf{x})}{\mathbf{f}'(\mathbf{x})}.$$

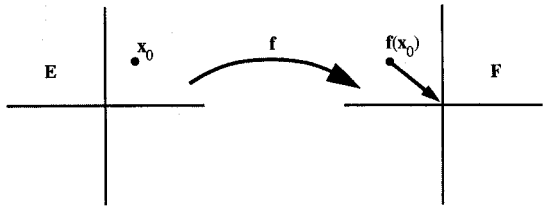

Figure 1

The system of equations **f** viewed as a transformation from one normed linear space to another. The straight arrow is the vector $-\mathbf{f}(\mathbf{x}_0)$.

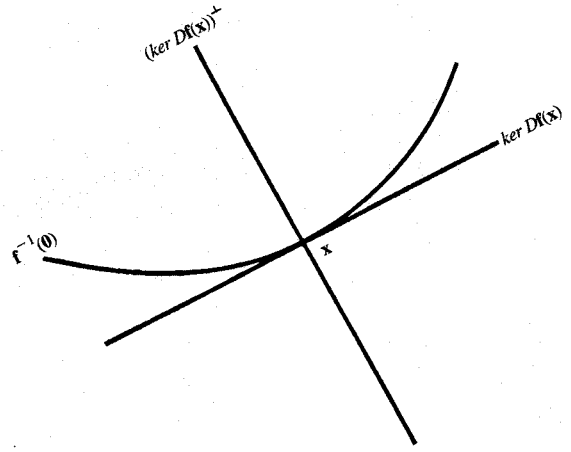

Figure 2

$\mathbf{f}^{-1}(\mathbf{0})$ as a graph of a function defined on *ker* $D\mathbf{f}(\mathbf{x})$.

Newton's method has the following two important features:

- Fixed points of $\mathbf{N}_f[\mathbf{N}_f(\xi) = \xi]$ correspond to zeros of **f** $[\mathbf{f}(\xi) = \mathbf{0}]$.
- At a simple zero of **f** $[\mathbf{f}(\xi) = \mathbf{0}$, and $D\mathbf{f}(\xi)$ is invertible], the derivative of Newton's method considered as a transformation, $D\mathbf{N}_f(\xi)$, is identically zero. Thus, the Taylor series of $\mathbf{N}_f$ at ξ begins with quadratic terms, and Newton's method converges quadratically to ξ in a ball around ξ.

We assume that **f** is twice continuously differentiable, C^2.

During the past three years, Steve Smale and the author have written a series of papers [1–4] about Newton's method and its extensions. In ths paper, some of the results are surveyed in three sections on the implicit function theorem, homotopies, and Bezout's theorem. In addition, a new example applying these methods to the theory of economic equilibria is given.

The implicit function theorem

If $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ with $m > n$, we may still solve the linearized equation $D\mathbf{f}(\mathbf{x}_0)\mathbf{v} = -\mathbf{f}(\mathbf{x}_0)$ for the case in which $D\mathbf{f}(\mathbf{x}_0)$ is onto [i.e., $D\mathbf{f}(\mathbf{x}_0)$ has rank n]. For this case, we use $D\mathbf{f}(\mathbf{x}_0)^{\dagger}$, the Moore–Penrose inverse of $D\mathbf{f}(\mathbf{x}_0)$, which is given by

$$D\mathbf{f}(\mathbf{x}_0)^{\dagger} = D\mathbf{f}(\mathbf{x}_0)^*(D\mathbf{f}(\mathbf{x}_0)D\mathbf{f}(\mathbf{x}_0)^*)^{-1},$$

where $D\mathbf{f}(\mathbf{x}_0)^*$ is the adjoint of $D\mathbf{f}(\mathbf{x}_0)$. It is easily seen that $D\mathbf{f}(\mathbf{x}_0)$ maps $\mathbb{R}^n$ to $\mathbb{R}^m$, its image is the orthogonal complement of the kernel of $D\mathbf{f}(\mathbf{x}_0)$, and it satisfies

$$D\mathbf{f}(\mathbf{x}_0)D\mathbf{f}(\mathbf{x}_0)^{\dagger} = Id_{\mathbb{R}^n}.$$

We now define Newton's method for $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$, where **f** is globally or locally defined, by

$$\mathbf{N}_f(\mathbf{x}) = \mathbf{x} - D\mathbf{f}(\mathbf{x})^{\dagger}\mathbf{f}(\mathbf{x}).$$

If $\mathbf{x}_0$ is a "starting point" in $\mathbb{R}^m$, we may define $\mathbf{x}_i = \mathbf{N}_f(\mathbf{x}_{i-1})$ as long as $D\mathbf{f}(\mathbf{x}_{i-1})$ is surjective. Note also that the hypothesis that $D\mathbf{f}(\mathbf{x})$ is surjective ensures that

Fixed points of $\mathbf{N}_f$ *correspond to zeros of* **f**.

This generalizes the first important feature of Newton's method. The generalization of the second important feature lies in the domain of the implicit function theorem. Let $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$, **f** be continuously differentiable, $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, and $D\mathbf{f}(\mathbf{x})$ be surjective. Then the implicit function theorem asserts that there is a neighborhood U of **x** such that $\mathbf{f}^{-1}(\mathbf{0}) \cap U$ is given as the graph of a C^1 function $\sigma : \mathbf{x} + \ker D\mathbf{f}(\mathbf{x}) \to \mathbf{x} + (\ker D\mathbf{f}(\mathbf{x}))^{\perp}$, where σ is defined on a neighborhood of zero in $\ker D\mathbf{f}(\mathbf{x})$, and $\ker D\mathbf{f}(\mathbf{x})$ is the null space of $D\mathbf{f}(\mathbf{x})$. (See **Figure 2**.)

In coordinates, we frequently have $\mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^n$, with $\mathbb{R}^k \equiv \ker D\mathbf{f}(\mathbf{x})$. Then $\sigma : \mathbf{x} + \mathbb{R}^k \to \mathbf{x} + \mathbb{R}^n$, where σ is defined near $\mathbf{0}$ in $\mathbb{R}^k$ and $\mathbf{f}(\mathbf{y}, \sigma(\mathbf{y})) = \mathbf{0}$. That is, σ is the implicit function. In the next proposition, taken from [4], **f** may be locally or globally defined, it is assumed to be of class C^2, and $\mathbf{0}$ is assumed to be a regular value of **f**; i.e., $D\mathbf{f}(\mathbf{x})$ is surjective for every **x** such that $\mathbf{f}(\mathbf{x}) = \mathbf{0}$.

Proposition 1 Suppose that $\mathbf{0}$ is a regular value of $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$. For $\zeta \in \mathbf{f}^{-1}(\mathbf{0})$, let

$W^s_\zeta = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{N}_\mathbf{f}^k(\mathbf{x}) \text{ converges to } \zeta \text{ as } k \to \infty\}$.

By $\mathbf{N}_\mathbf{f}^k$, we mean the kth iterate of $\mathbf{N}_\mathbf{f}$. Then

- The union of W^s_ζ over $\zeta \in \mathbf{f}^{-1}(\mathbf{0})$ is a neighborhood of $\mathbf{f}^{-1}(\mathbf{0})$.
- The intersection of W^s_ζ and a small neighborhood of $\mathbf{f}^{-1}(\mathbf{0})$ is a cell varying continuously in ζ.
- $D\mathbf{N}_\mathbf{f}(\zeta)$ restricted to $(ker\ D\mathbf{f}(\zeta))^\perp$ is zero. The tangent space of W^s_ζ at ζ is the orthogonal complement to $T_\zeta(\mathbf{f}^{-1}(\mathbf{0})) = (ker\ D\mathbf{f}(\zeta))^\perp$.

This extends the usual basin of attraction theory from the case $m = n$. The W^s_ζ are illustrated as fibers in **Figure 3**.

To obtain more information on the size of the neighborhoods and speed of convergence, we might exploit C^2 estimates in a neighborhood of $\mathbf{f}^{-1}(\mathbf{0})$, or higher-order estimates along $\mathbf{f}^{-1}(\mathbf{0})$ itself. We take the latter approach; henceforth we assume that $\mathbf{f}$ is real analytic. Define for $\mathbf{x} \in \mathbb{R}^m$

$$\beta(\mathbf{f}, \mathbf{x}) \equiv \|D\mathbf{f}(\mathbf{x})^\dagger \mathbf{f}(\mathbf{x})\| \quad [\text{or } \infty \text{ if } D\mathbf{f}(\mathbf{x}) \text{ is not surjective}],$$

$$\gamma(\mathbf{f}, \mathbf{x}) \equiv \max_{k>1} \left(\left\| D\mathbf{f}(\mathbf{x})^\dagger \frac{D^k\mathbf{f}(\mathbf{x})}{k!} \right\| \right)^{1/(k-1)}$$

[or ∞ if $D\mathbf{f}(\mathbf{x})$ is not surjective],

and

$$\alpha(\mathbf{f}, \mathbf{x}) \equiv \beta(\mathbf{f}, \mathbf{x})\gamma(\mathbf{f}, \mathbf{x}).$$

Theorem 1 There is a universal constant α_0, approximately 1/7, such that if $\mathbf{f}$ and $\mathbf{x}$ are as above, with $\mathbf{a}(\mathbf{f}, \mathbf{x}) < \alpha_0$ and $\mathbf{x} = \mathbf{x}_0$, then (a) all the Newton iterates $\mathbf{x}_1, \mathbf{x}_2, \cdots$ are defined and converge to $\zeta \in \mathbb{R}^n$, with $\mathbf{f}(\zeta) = \mathbf{0}$, and (b) for all $k \geq 1$,

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \leq \left(\frac{1}{2}\right)^{2^{k-1}} \|\mathbf{x}_1 - \mathbf{x}_0\|. \tag{2}$$

A point $\mathbf{x}_0 \in \mathbb{R}^m$ is called an *approximate zero* of $\mathbf{f}$ if (2) is satisfied. Then ζ is called the associated zero.

We also verify, in the following theorem from [4], that a point is an approximate zero in terms of γ along $\mathbf{f}^{-1}(\mathbf{0})$ and the distance to $\mathbf{f}^{-1}(\mathbf{0})$.

Theorem 2 Let $\mathbf{F} : \mathbb{R}^m \to \mathbb{R}^n$ have zero as a regular value, and define $\gamma \equiv \max_{\mathbf{z} \in \mathbf{f}^{-1}(\mathbf{0})} \gamma(\mathbf{F}, \mathbf{z})$. Then there is a universal constant C such that if the distance $d(\mathbf{z}', \mathbf{F}^{-1}(\mathbf{0})) < C/\gamma$, then $\mathbf{z}'$ is an approximate zero.

Remark Let $\mathbf{f} = (f_1, \cdots, f_n)$ and let each f_i be homogeneous of some degree d_i; i.e.,

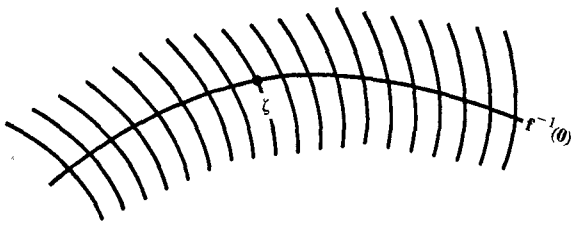

Figure 3

The *fibering* of a neighborhood of $\mathbf{f}^{-1}(\mathbf{0})$ by the cells of Proposition 1.

$$\mathbf{f}(\lambda\mathbf{x}) = (f_1(\lambda\mathbf{x}), \cdots, f_n(\lambda\mathbf{x})) = (\lambda^{d_1} f_1(\mathbf{x}), \cdots, \lambda^{d_n} f_n(\mathbf{x})).$$

It is then easy to see, using the chain rule, that $\beta(\mathbf{f}, \lambda\mathbf{x}) = \lambda\beta(\mathbf{f}, \mathbf{x})$ and $\gamma(\mathbf{f}, \lambda\mathbf{x}) = (1/\lambda)\gamma(\mathbf{f}, \mathbf{x})$. Thus $\alpha(\mathbf{f}, \lambda\mathbf{x}) = \alpha(\mathbf{f}, \mathbf{x})$ for $\lambda \neq 0$.

Example 1 Homogeneous functions naturally arise in economic theory. Given ℓ commodities and prices $\mathbf{p} = (p_1, \cdots, p_\ell)$, where $p_i > 0$ for $i = 1, \cdots, \ell$, one may define the excess-demand function $\mathbf{f}(\mathbf{p}) = D(\mathbf{p}) - \mathbf{S}(\mathbf{p})[\mathbf{f}(\mathbf{p}) \in \mathbb{R}^\ell]$. Here $D(\mathbf{p})$ is the demand for the ℓ commodities, and $\mathbf{S}(\mathbf{p})$ is the supply at prices $\mathbf{p}$. Thus $\mathbf{f}$ maps the positive orthant $\mathbb{R}^\ell_+$ to the ℓ-dimensional commodity space. An equilibrium is given by supply equals demand; i.e., $\mathbf{f}(\mathbf{p}) = \mathbf{0}$. There are two additional conditions imposed on $\mathbf{f}$:

- (Walras' law) The dot product $\mathbf{p} \cdot \mathbf{f}(\mathbf{p}) = 0$, which reflects the hypothesis that each economic agent can demand only goods whose value equals the value of his supply.
- $\mathbf{f}(\lambda\mathbf{p}) = \mathbf{f}(\mathbf{p})$ for $\lambda > 0$, which states that supply and demand depend only on relative prices, not on the units chosen to express them.

Let $\mathbf{E} \subset \mathbb{R}^\ell$ be the subspace $\{\mathbf{x} \in \mathbb{R}^\ell \mid \Sigma x_i = 0\}$. Then $\mathbf{E}$ has dimension $\ell - 1$. Let $\mathbf{g} : \mathbb{R}^\ell_+ \to \mathbf{E}$ be defined by $\mathbf{g}(\mathbf{p}) \equiv \mathbf{f}(\mathbf{p}) - (\Sigma\mathbf{f}(\mathbf{p})_i/\Sigma p_i)\mathbf{p}$. Then it is quite easy to see that $\mathbf{g}(\mathbf{p}) = \mathbf{0}$ if and only if $\mathbf{f}(\mathbf{p}) = \mathbf{0}$. The proof is as follows: If $\mathbf{f}(\mathbf{p}) = \mathbf{0}$, then $\mathbf{f}(\mathbf{p})_i = \mathbf{0}$ for each i, and $\mathbf{g}(\mathbf{p}) = \mathbf{0}$. On the other hand, if $\mathbf{g}(\mathbf{p}) = \mathbf{0}$, then $\mathbf{f}(\mathbf{p})$ is a scalar multiple of $\mathbf{p}$ (say $\mu\mathbf{p}$); however, by Walras' law, $\mathbf{p} \cdot \mathbf{f}(\mathbf{p}) = \mathbf{p} \cdot \mu\mathbf{p} = \mu\|\mathbf{p}\|^2 = 0$, so $\mu = 0$ and $\mathbf{f}(\mathbf{p}) = \mathbf{0}$. It is straightforward to show that the image of $\mathbf{g}$ lies in $\mathbf{E}$. In this way, we may interpret the problem of finding the supply-equals-demand equilibrium as the problem of finding the zeros of the homogeneous function $\mathbf{g} : \mathbb{R}^\ell_+ \to \mathbf{E}$.

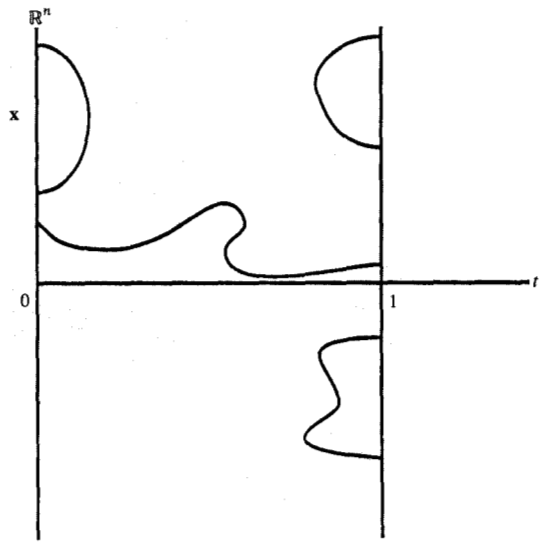

Figure 4

$\mathbf{F}^{-1}(\mathbf{0})$ as seen in $I \times \mathbb{R}^n$.

Homotopies

In this section, we assume that we can replace an approximate zero with the precise zero to which it is converging. One can imagine adding such an operation or a node to the B–S–S model of computation over the reals [5]. This hypothesis seems reasonable not only because of the rapid convergence of the approximate zero to the precise zero but also by the extensive analysis done in the approximate case in [1]. For what we say below, we may use approximate zeros at only slightly greater computational cost (perhaps a factor of three).

One method for finding zeros of functions $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is to start with a function whose zeros are known, $\mathbf{f}_0$, and to produce the homotopy $\mathbf{f}_t = (1 - t)\mathbf{f}_0 + t\mathbf{f}$ $(0 \le t \le 1)$, so that $\mathbf{f}_1 = \mathbf{f}$. Let I be the unit interval. If we assume that 0 is a regular value of the joint map $\mathbf{F} : I \times \mathbb{R}^n \to \mathbb{R}^n$, where $\mathbf{F}(t, \mathbf{x}) = \mathbf{f}_t(\mathbf{x})$, then $\mathbf{F}^{-1}(\mathbf{0})$ is the union of a finite number of paths. (See **Figure 4**.)

If we assume, moreover, that $\mathbf{F}$ is proper, these paths cannot run off to infinity. Following these paths may lead us from the zeros of $\mathbf{f}_0$ to some of the zeros of $\mathbf{f}_1$. Allgower and Georg [6] propose an algorithm for following an arc A of $\mathbf{F}^{-1}(\mathbf{0})$ for $\mathbf{F} : \mathbb{R}^{n+1} \to \mathbb{R}^n$, with 0 as a regular value. The algorithm moves tangentially to the curve in a predictor step and then uses the Moore–Penrose Newton method as a corrector. We assume $\mathbf{F}$ to be real analytic and make sure that the predictor step gives an approximate zero for $\mathbf{F}$, which we then replace with its associated zero. We call this a predictor–corrector step. From [4] we have Theorem 3.

Theorem 3 The complexity (number of predictor–corrector steps) sufficient to follow an arc A of $\mathbf{F}^{-1}(\mathbf{0})$ (where $\mathbf{F} : \mathbb{R}^{n+1} \to \mathbb{R}^n$, as above) is $C\gamma L$, where L is the length of A, C is a constant (not more than 20), and $\gamma = \max_{\mathbf{x} \in A} \gamma(\mathbf{F}, \mathbf{x})$.

Theorem 1 can be used quite generally to find an upper bound on the complexity of following homotopies. The next result is once again from [4]. Consider $\mathbf{f}_t : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{y}_t \in \mathbb{R}^n$, a homotopy and path respectively for $0 \le t \le 1$, and let $\boldsymbol{\zeta}_0 \in \mathbb{R}^n$ satisfy $\mathbf{f}_0(\boldsymbol{\zeta}_0) = \mathbf{y}_0$. Define

$$A_{t,t'} = \max_{\substack{\mathbf{x} \text{ subject to} \\ \mathbf{f}_t(\mathbf{x}) = \mathbf{y}_t}} \alpha(\mathbf{f}_{t'} - \mathbf{y}_{t'}, \mathbf{x}).$$

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

Hypothesis Suppose that $A_{t,t'} < \alpha_0$ whenever $|t - t'| \le \Delta \equiv 1/k$, where k is a positive integer.

Corollary of Theorem 1 Let $\mathbf{f}_t$, $\mathbf{y}_t$, $\boldsymbol{\zeta}_0$ be as above and satisfy the hypothesis. Then k of the hypothesis is a sufficient number of steps to solve $\mathbf{f}_1(\boldsymbol{\zeta}_1) = \mathbf{y}_1$.

The proof from [4] is so simple we repeat it here. Let $t_0 = 0$ and $t_i = t_{i-1} + \Delta$; thus, $\alpha(\mathbf{f}_{t_i}, \boldsymbol{\zeta}_{t_{i-1}}) < \alpha_0$. Then a Moore–Penrose Newton step (corrected to yield the associated zero) yields $\boldsymbol{\zeta}_{t_i}$ from $\boldsymbol{\zeta}_{t_{i-1}}$, starting from $\boldsymbol{\zeta}_0$, with $\mathbf{f}(\boldsymbol{\zeta}_{t_i}) = \mathbf{y}_{t_i}$. The following two examples are natural candidates for application of the corollary.

Example 2 Let $\mathbf{f} : \mathbb{R}^m \to R^n$, $\mathbf{f}_t \equiv \mathbf{f}$, $\mathbf{f}(\mathbf{x}_0) = \mathbf{y}_0$, and $\mathbf{y}_t = t\mathbf{f}(\mathbf{x}_0)$. In this version, see [7] for the case $m = n$. The resulting algorithm has been extensively studied for univariate polynomials over the complex numbers. See [7, 8] and the references therein.

Example 3 Let $\mathbf{g} : \mathbb{R}_+^\epsilon \to \mathbf{E}$ be defined as in Example 1. Let $\mathbf{g}_t \equiv \mathbf{g}$ and $\mathbf{y}_t \equiv t\mathbf{g}(\mathbf{x}_0)$. This is a version of Smale's global Newton's equation (see [9]), for which we now have a complexity estimate in terms of A.

For more examples, applications, and discussions, see [4].

Bezout's theorem

Bezout's theorem is the higher-dimensional analogue of the fundamental theorem of algebra. The fundamental theorem of algebra asserts that a complex polynomial of degree d, $p(z) = a_d z^d + \cdots + a_0$ (where $a_i \in \mathbb{C}$ the complex numbers), has d complex roots. There are two provisos:

- $a_d \neq 0$.
- The roots must be counted with multiplicity.

The first proviso of this theorem may be eliminated by considering homogeneous equations. Let $\hat{P}(z, w) = a_d z^d + a_{d-1} z^{d-1} w + a_0 w^d$ be a homogeneous complex polynomial. Roots are in $\mathbb{C}^2$, and since $\hat{P}(\lambda z, \lambda w) = \lambda^d \hat{P}(z, w)$ for $\lambda \in \mathbb{C}$, the roots consist of whole complex lines in $\mathbb{C}^2$. Thus, the fundamental theorem of algebra now asserts that $\hat{P}(z, w)$ has d solution lines in $\mathbb{C}^2$, with the sole proviso that they be counted with multiplicity.

Let $f_i : \mathbb{C}^{n+1} \to \mathbb{C}$ be a homogeneous complex polynomial of degree d_i, for $i = 1, \cdots, n$. Bezout's theorem asserts that the system of equations

$$f_1(\mathbf{x}) = 0, \cdots, f_n(\mathbf{x}) = 0$$

has $\mathcal{D} = \Pi_{i=1}^n d_i$ solution lines in $\mathbb{C}^{n+1}$, counted with multiplicity. The multiplicity is 1 if $D\mathbf{f}(\mathbf{x})$ has rank n at the solution $\mathbf{x}$, [$\mathbf{f} = (f_1, \cdots, f_n)$, $\mathbf{f} : \mathbb{C}^{n+1} \to \mathbb{C}^n$]. It is generally the case that the multiplicity is 1 for all of the solutions.

In [1–4], we have investigated homotopy methods to find all of the roots of $\mathbf{f}$ given all of the roots of another system $\mathbf{g}$. We use a projective Newton method suggested in [10] instead of Moore–Penrose. Let $\langle , \rangle$ be the standard Hermitian product on $\mathbb{C}^{n+1}$ [$\langle \mathbf{x}, \mathbf{y} \rangle = \Sigma_i^{n+1} x_i \bar{y}_i$]. Let $null\ \mathbf{x} \equiv \{\mathbf{v} \in \mathbb{C}^{n+1} \mid \langle \mathbf{v}, \mathbf{x} \rangle = 0\}$. Then the projective Newton method is defined by

$$\mathbf{N}_\mathbf{f}(\mathbf{x}) = \mathbf{x} - (D\mathbf{f}|_{null\ \mathbf{x}})(\mathbf{x})^{-1}\mathbf{f}(\mathbf{x}).$$

Thus, the image of $(D\mathbf{f}|_{null})(\mathbf{x})^{-1}$ is the orthogonal complement to the line through $\mathbf{x}$ in $\mathbb{C}^{n+1}$. If $\boldsymbol{\xi}$ is a nondegenerate root of $\mathbf{f}$, then $(ker\ D\mathbf{f}(\boldsymbol{\xi}))^\perp = null\ \boldsymbol{\xi}$; however, this is not the case in general, so projective Newton and Moore–Penrose Newton differ.

Let $\mathcal{H}_{(\mathbf{d})}$, $(\mathbf{d}) = (d_1, \cdots, d_n)$ be the complex vector space of systems of homogeneous polynomial equations $\mathbf{f} = (f_1, \cdots, f_n)$, where $f_i : \mathbb{C}^{n+1} \to \mathbb{C}^1$ is a homogeneous polynomial of degree d_i.

The algorithm proposed in [1] considers the homotopy $\mathbf{f}_t = (1 - t)\mathbf{g} + t\mathbf{f}$, where $0 \leq t \leq 1$. One of our results, proven in [4], gives a bound on the number of projective Newton steps required to find all of the approximate zeros of $\mathbf{f}$ (i.e., one approximate zero corresponding to each precise root). The bound depends only on $(\mathbf{d})$ and the probability of success σ.

Theorem 4 The number of projective Newton steps sufficient to find all of the approximate zeros of $\mathbf{f} \in \mathcal{H}_{(\mathbf{d})}$ with probability σ of success is

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

where $D = \max_i(d_i)$, $\mathcal{D} = \Pi_{i=1}^n d_i$, N is the dimension of $\mathcal{H}_{(\mathbf{d})}$, and c is a constant independent of n, $(\mathbf{d})$, and σ.

The space $\mathcal{H}_{(\mathbf{d})}$ is given a natural unitarily invariant Hermitian product. The measure on $\mathcal{H}_{(\mathbf{d})}$ is the usual Gaussian distribution given by the Hermitian structure.

Remarks While we know that there exists a g for each $\mathcal{H}_{(\mathbf{d})}$, we do not know how to find it, even for $n = 1$! See [2] for a discussion of this.

For $n = 1$, the number of steps is $cd^6/(1 - \sigma)$.

Reference [11] is an important predecessor to this paper. Specialized to $n = 1$, Renegar's result has a factor $d^{26}/(1 - \sigma)^4$. In [12] (which applies to only one variable), there is a similar result, with $d^9/(1 - \sigma)^7$.

Experiments by Raymond Russell (a student at Trinity College, Dublin), carried out at Berkeley, seem to support these findings.* For polynomials in one variable, homogenizing the homotopy and using projective Newton in place of Newton for the nonhomogenized homotopy produced significant speedup. For nonpolynomial systems, speedup may be achieved by homogenizing to degree 0 and using Moore–Penrose or projective Newton. Much more experimentation is called for here.

For one variable, more efficient algorithms are known for finding all of the roots. See [4] for references.

The number of steps in Theorem 4 must be interpreted as parallel steps. The algorithm moves along all $\mathcal{D}$ paths of roots simultaneously. For each path, the number of steps determined in Theorem 4 is required. For finding one root at a time, we think it likely that the upper bound in Theorem 4 could be divided by $\mathcal{D}$, giving a $\mathcal{D}^2$ factor for total speedup.

References

1. M. Shub and S. Smale, "Complexity of Bezout's Theorem I: Geometric Aspects," *J. Amer. Math. Soc.* **6,** 459–501 (1993).
2. M. Shub and S. Smale, "Complexity of Bezout's Theorem II: Volumes and Probabilities," *Computational Algebraic Geometry*, F. Eyssette and A. Galligo, Eds., *Progress in Mathematics*, Vol. 109, Birkhauser, Boston, 1993, pp. 267–285.
3. M. Shub and S. Smale, "Complexity of Bezout's Theorem III: Condition Number and Packing," *J. Complexity* **9,** 4–14 (1993).
4. M. Shub and S. Smale, "Complexity of Bezout's Theorem IV: Probability of Success, Extensions," *Research Report 18921*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1994.
5. L. Blum, M. Shub, and S. Smale, "On a Theory of Computation and Complexity Over the Real Numbers: NP-Completeness, Recursive Functions, and Universal Machines," *Bull. Amer. Math. Soc.* **21,** 1–46 (1989).

*Steve Smale, Department of Mathematics, University of California at Berkeley, March 21, 1994, personal communication.

6. E. Allgower and K. Georg, "Continuation and Path Following," *Acta Numer.*, pp. 1–64 (1993).
7. S. Smale, "Algorithms for Solving Equations," *Proceedings of the International Congress of Mathematics*, Berkeley, 1986, Amer. Math. Soc., Providence, RI, 1987.
8. S. Smale, "Newton's Method Estimates from Data at One Point," *The Merging of Disciplines: New Directions in Pure, Applied and Computation Math*, R. Ewing, K. Gross, and C. Martin, Eds., Springer, New York, 1986.
9. S. Smale, "Global Analysis and Economics," *Handbook of Mathematical Economics*, Vol. 1, K. Arrow and M. Intrilligator, Eds., North Holland Publishing Co., New York, 1981, Ch. 8.
10. M. Shub, "Some Remarks on Bezout's Theorem and Complexity Theory," *From Topology to Computation, Proceedings of the Smalefest*, M. Hirsch, J. Marsden, and M. Shub, Eds., Springer-Verlag, New York, 1993.
11. J. Renegar, "On the Efficiency of Newton's Method in Approximating All Zeros of Systems of Complex Polynomials," *Math. Oper. Res.* **12,** 121–148 (1987).
12. S. Smale, "The Fundamental Theorem of Algebra and Complexity Theory," *Bull. Amer. Math. Soc. (N.S.)* **4,** 1–36 (1981).

Received June 11, 1993; accepted for publication February 4, 1994

Michael Shub *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (SHUB at YKTVMV, shub@watson.ibm.com).* Dr. Shub received his Ph.D in mathematics in 1967 from the University of California at Berkeley. He has been a Research Staff Member in the Mathematical Sciences Department of the IBM Thomas J. Watson Research Center since 1986. In the last 27 years he has authored and coauthored 82 publications on chaos and complexity.