Numer. Math. 16, 334–342 (1971)

Deflation Techniques for the Calculation of Further Solutions of a Nonlinear System

Kenneth M. Brown and William B. Gearhart

Received March 10, 1970

Summary. This paper defines several classes of methods which can be used to find additional solutions of a nonlinear system of equations. A theory which embraces these classes is presented and the theory is extended to the multiple root problem. The techniques developed can also be used in avoiding previously found extreme points when performing function minimization. Results of computer experiments are presented.

1. Introduction

We consider a system of N nonlinear real equations in N real unknowns. There are a variety of numerical methods for obtaining solutions of such systems. This paper presents the theory and applications for classes of methods which can be used to find *further* solutions—in addition to those known a priori or found during earlier calculations. The basis for these methods is the following technique: once a root has been determined, a new system of equations is formed in such a way that it retains those roots of the original system which remain to be computed, but no longer tends to zero values as the old root is approached. This technique is called *deflation* and has been studied in the one dimensional case by, e.g., Wilkinson [4, pp. 55–65, 78].

After introducing the concept of a *deflation matrix*, a theory of deflation for nonlinear systems is developed and extended to include the multiple root problem. Some specific deflation techniques are given and their application to function minimization is shown. Computer results are presented in which the techniques developed are applied to specific problems. Of interest is a problem possessing a *magnetic zero*, a zero converged to almost regardless of the starting guess used. These zeros often mask out the zeros of real interest. The techniques introduced have particular merit in enabling one to avoid convergence to such magnetic zeros.

2. Deflation Matrices

Let $f(x)=0$ denote the vector form of a system of N real nonlinear equations in N real unknowns. Let E^N denote N-dimensional Euclidean space, with the inner product of $u, v \in E^N$ denoted by $\langle u, v\rangle$, and the associated norm by $\|u\|$. For each $r \in E^N$, let $M(x;r)$ be a matrix on E^N which is defined for all $x \in U_r$, where U_r is some open set in E^N such that r belongs to the closure of U_r. We will say that M is a *deflation matrix* if for any differentiable $f\colon E^N \to E^N$ such that $f(r)=0$ and $f'(r)$ is nonsingular, we have

$$\liminf_{i\to\infty}\|M(x_i;r)\,f(x_i)\|>0 \tag{1}$$

for any sequence $x_i\to r$, $x_i\in U_r$. Here f' denotes the first Frechet derivative of f. We remark that it was not necessary to require that f be defined on all of E^N; however, we have done so for convenience.

Let us now form the function

$$F(x)=M(x;r)\,f(x); \tag{2}$$

thus, in applying an iterative method to find further zeros of F, we are assured—in the sense described by (1)—that any sequence of points converging to r (a simple zero of f) will not produce a zero of F. We shall call F the *deflated function*, or the function obtained from f by *deflating out* the simple zero $x=r$.

To deflate out k simple roots, $r_1, r_2, \ldots, r_k$, we form the deflated function

$$F(x)=M(x;r_1)\,M(x;r_2)\ldots M(x;r_k)\,f(x). \tag{3}$$

The following result is helpful in constructing deflation matrices.

Lemma 2.1. Suppose the matrix $M(x;r)$ has the property:

(P) For each $r\in E^N$ and any sequence

$$x_i\to r,\quad x_i\in U_r,\quad \text{if}\quad \|x_i-r\|\,M(x_i;r)\,u_i\to 0$$

for some sequence $\{u_i\}\subset E^N$, then $u_i\to 0$.

Then M is a deflation matrix.

Proof. If not, then there is a differentiable function $f\colon E^N\to E^N$ and an $r\in E^N$ such that $f(r)=0$, $f'(r)$ is nonsingular, and

$$\liminf_{i\to\infty}\|M(x_i;r)\,f(x_i)\|=0$$

for some sequence $x_i\to r$, $x_i\in U_r$. But then there is a subsequence, which we again denote by $\{x_i\}$, such that $M(x_i;r)\,f(x_i)\to 0$, and therefore

$$\|x_i-r\|\,M(x_i;r)\,u_i\to 0$$

where we have set $u_i=f(x_i)/\|x_i-r\|$. From property (P), we may conclude that $f(x_i)/\|x_i-r\|\to 0$. But then it would follow that $f'(r)\cdot(x_i-r)/\|x_i-r\|\to 0$, which contradicts the nonsingularity of $f'(r)$.

3. Some Classes of Deflations for Nonlinear Systems

In this section we consider two classes of deflations for nonlinear systems. Using Lemma 2.1 it is not difficult to show that the matrices associated with these classes are deflation matrices.

I. Norm Deflation. In (3) we take $M(x;r_i)=\frac{1}{\|x-r_i\|}A$, where A is a nonsingular matrix on E^N. Here, any vector norm can be employed. The domain of definition for (3) is given by $E^N-\bigcup_{i=1}^{k}\{r_i\}$. In the numerical results reported in Section 5, we have used $A=I$, the identity matrix; sagacious choices of A are being investigated for purposes of accelerating convergence or preconditioning the Jacobian matrix of the system at certain points. Forsythe and Moler [3, p. 136] have also proposed this special form of norm deflation in which $A=I$.

II. Inner Product Deflation. Here we take $M(x; r_i)$ in (3) to be a diagonal matrix whose j-th diagonal element is given by

$$m_{jj} = (\langle a_j^{(i)}, x - r_i \rangle)^{-1}$$

where $a_1^{(i)}, \ldots, a_n^{(i)}$ are nonzero vectors in E^N, $i=1, \ldots, k$. If we set

$$C(a_j) = \{u \in E^N \colon \langle a_j^{(i)}, u \rangle = 0,\ i=1, \ldots, k\}$$

then the domain of definition for (3) is given by

$$E^N - \bigcup_{i=1}^{k} \bigcup_{j=1}^{N} \left(C(a_j^{(i)}) + \{r_i\}\right).$$

If the $a_j^{(i)}$ are chosen as

$$a_j^{(i)} = \operatorname{grad} f_j|_{x=r_i} \equiv f_j'(r_i)$$

we shall refer to the method as *gradient deflation.* (This choice has proven useful in practice when using Newton's Method.) The j-th component of F in (3) for gradient deflation is given by

$$F_j(x) = \frac{f_j(x)}{\prod\limits_{i=1}^{k} \langle \operatorname{grad} f_j(r_i), x - r_i \rangle}, \qquad j=1, \ldots, N. \tag{4}$$

Remark 3.1. The deflation techniques presented here for E^N have the advantage of keeping the iterates away from previously determined roots; however, unlike the one dimensional case, there is no way to define the deflated function F for all points at which f is defined. Possible choices for $F(r)$ (relative to particular iterative methods) which aid, or at least do not hinder, convergence to other roots are being investigated.

Remark 3.2. Should (4) be too unwieldy for the practical computations necessitated by a particular problem, it may be more convenient to use a finite difference approximation to any partial derivative of F required by the iterative technique being used; e.g.,

$$\frac{\partial F_i}{\partial x_j} \sim \frac{F_i(x + h e_j) - F_i(x)}{h}$$

where F_i denotes the i-th component function of the vector function F, where e_j is the j-th unit vector and h is a small positive increment which may depend on x. In the numerical results reported in Section 5, we have used these first differences to approximate the partials needed.

4. Multiple Zeros

The definition of the deflation matrix indicates the behavior of the matrix with regard to only simple zeros. We will consider now the effect of the deflation matrix on multiple zeros of $f \colon E^N \to E^N$. For convenience, we will assume that f is infinitely differentiable and let $f^{(k)}$ denote the k-th Frechet derivative of f

(thus, $f^{(k)}(x) \cdot v^{(k)}$ denotes the k-th derivative of f at $x \in E^N$ applied k times to the vector $v \in E^N$).

If there is a unit vector $v \in E^N$ such that $f^{(h)}(r) \cdot v^{(h)} = 0$ for all $h = 0, 1, 2, \ldots$, then we will say that f is *flat* at $x = r$.

In E^1 it is easy to show that if we apply the deflation $1/(x-r)$ to $f\colon E^1 \to E^1$, then provided f is not flat at $x = r$, there is an integer k such that $f(x)/(x-r)^k$ does not have a zero at $x = r$. In performing the numerical calculations, it is not necessary to know the multiplicity of a given zero in order to construct the deflated function. Indeed, when a zero, say $x = r$, is first computed, it suffices to merely divide by the factor $(x-r)$. Each time this zero is computed again (if at all), we simply divide again by the factor $(x-r)$. Thus, unless f is flat at $x = r$, we will ultimately arrive at a deflated function which does not have a zero at r. In higher dimensions, we have

Proposition 4.1. Suppose $f\colon E^N \to E^N$ is infinitely differentiable and $f(r) = 0$. Let $M(x; r)$ have property (P) in Lemma 2.1 (in particular, M is a deflation matrix). Then, provided f is not flat at $x = r$, there is an integer K such that

$$\liminf_{i \to \infty} \|M^K(x_i; r) f(x_i)\| > 0$$

for any sequence $x_i \to r$, $x_i \in U_r$.

Proof. If the conclusion is false, then for each integer $k \geqq 1$, there corresponds a sequence $\{x_i^{(k)}\}_{i \geqq 1} \subset E^N$ such that $x_i^{(k)} \to r$ and $M^k(x_i^{(k)}; r) f(x_i^{(k)}) \to 0$ as $i \to \infty$. But then

$$\left(\|x_i^{(k)} - r\| M(x_i^{(k)}; r)\right)^k \cdot u_i^{(k)} \to 0$$

where $u_i^{(k)} = f(x_i^{(k)})/\|x_i^{(k)} - r\|^k$, and using property (P), it follows that $f(x_i^{(k)})/\|x_i^{(k)} - r\|^k \to 0$. Expanding f in a Taylor series, we have

$$f(x_i^{(k)}) = \sum_{h=0}^{k-1} \frac{1}{h!} f^{(h)}(r) \cdot (x_i^{(k)} - r)^{(h)} + \frac{1}{k!} f^{(k)}(\xi) \cdot (x_i^{(k)} - r)^{(k)}$$

so that

$$\frac{f(x_i^{(k)})}{\|x_i^{(k)} - r\|^k} = \sum_{h=1}^{k-1} \frac{1}{h!} \frac{f^{(h)}(r) \cdot (v_i^{(k)})^{(h)}}{\|x_i^{(k)} - r\|^{k-h}} + \frac{1}{k!} f^{(k)}(\xi) \cdot (v_i^{(k)})^{(k)} \tag{5}$$

where $v_i^{(k)} = (x_i^{(k)} - r)/\|x_i^{(k)} - r\|$. Since $\|v_i^{(k)}\| = 1$ for all i, we may assume that $v_i^{(k)} \to v_k$ where $\|v_k\| = 1$. Now, if we multiply Eq. (5) by $\|x_i^{(k)} - r\|^{k-1}$ and take the limit as $i \to \infty$, then it follows that $f^{(1)}(r) \cdot (v_k)^{(1)} = 0$. If we next multiply Eq. (5) by $\|x_i^{(k)} - r\|^{k-2}$ and let $i \to \infty$, then we see that $f^{(2)}(r) \cdot (v_k)^{(2)} = 0$. Continuing in this fashion, we conclude that $f^{(h)}(r) \cdot (v_k)^{(h)} = 0$ for $h = 1, 2, \ldots, k$. Now consider the sequence $\{v_k\}_{k \geqq 0}$. Since $\|v_k\| = 1$ for all k, we may assume that $v_k \to v$ where $\|v\| = 1$; therefore, $f^{(h)}(r) \cdot v^{(h)} = 0$ for all $h = 0, 1, 2, \ldots$, which is a contradiction.

From this proposition, we see that a deflation matrix with property (P) treats multiple zeros in a fashion similar to that of the deflation $1/(x-r)$ for functions of one variable. If we assume in addition that the deflation matrix is such that $M(x; r)$ and $M(x; s)$ commute for each $r, s \in E^N$, then in the numerical calculations, we can deflate out multiple zeros using the same procedure as was outlined for the deflation $1/(x-r)$ in E^1. Clearly both norm deflation and inner product deflation have this commutative property.

5. Numerical Results

Two basic iterative methods of solution in conjunction with three types of deflation were tested on various nonlinear systems. The iterative methods used were

1) the discretized form of Newton's Method and

2) a discretized form [1] of a quadratically convergent method due to Brown [2].

The deflation techniques were

1) norm deflation with the uniform norm,

2) norm deflation with the Euclidean norm, and

3) gradient deflation;

deflations 1) and 2) will hereafter be referred to as l_∞ deflation and l_2 deflation, respectively. The strategy of all computations was as follows: given a nonlinear system and a starting guess x_0 one of the solution techniques (discretized Newton's or Brown's) was used to locate a first zero. When accomplished one of the deflation techniques was then used with the same method beginning with the *same* starting guess x_0 to find a second zero. If a second zero was found, deflation was performed again, and an attempt was made to find a third zero. This process was continued, always starting with the same x_0, until:

(a) all zeros of the system (or the maximum number requested by the programmer) had been found;

(b) the process diverged to "infinity";

(c) the maximum number of iterations allowed was exceeded; or

(d) the iterates reached a point at which the Jacobian matrix became singular.

All of the test cases were run in FØRTRAN IV using double precision arithmetic on the IBM 360/65 at the Office of Computer Services of Cornell University. The convergence criterion was usually set to obtain twelve significant digits of accuracy.

Example 1. The Cubic-Parabola (see Fig. 1)

Consider the nonlinear system given by

$$f(x, y) = 4x^3 - 3x - y; \quad g(x, y) = x^2 - y.$$

This system has the three zeros $r_1 = (1, 1)$, $r_2 = (0, 0)$ and $r_3 = (-0.75, 0.5625)$. An interesting feature of this system is the fact that relative to the discretized Newton's Method, r_1 seems to be a *magnetic zero* which we define as a zero to which a particular method seems to converge no matter what the starting guess happens to be. We gave the discretized Newton's Method 21 starting guesses in the square bounded by $x = \pm 1$, $y = \pm 1$. None of the starting guesses was closer to the first zero than to any of the other zeros, yet 20 out of the 21 starting

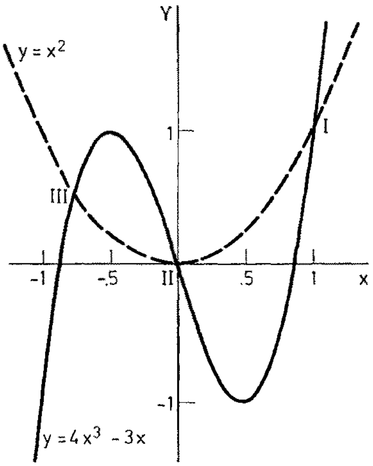

Fig. 1. Zero curves of the cubic-parabola system

guesses converged to r_1! (One cannot help but feel sorry for the harried scientist whose physical intuition tells him that there must be another solution, yet whose computer printout always meanders back to r_1 for a seemingly all encompassing set of starting guesses.) The presence of such a magnetic zero can often mask out the zeros of real interest. After removing this zero by deflation using Newton's Method, r_2 was found in about half the cases tested, the other half yielding divergence. (The results for the three types of deflation were qualitatively the same.) In no case was r_3 found with Newton's Method even though guesses correct to one significant digit were used. On the other hand using Brown's method r_3 *was* found for some of the 21 starting guesses; this indicates a difference in the regions of convergence of the two methods and shows that the "magnetism" of a particular zero is a function of the method used as well as the system. For several starting guesses, one being $x_0 = (0.8, 0.55)$, Brown's method with l_∞ deflation found *all* three zeros of the system (in the order r_3, r_2, r_1). Beginning with the same starting guess, the zeros were found in the order r_3, r_2, D using l_2 deflation, and in the order r_3, r_1, D using gradient deflation, where D stands for divergence. Newton's Method with this x_0 produced the sequence r_1, r_2, D for all three deflation techniques. The number of iterations used per zero by each method was comparable (12 or fewer).

Example 2. The Four-Cluster (see Fig. 2)

The system

$$f(x, y) = (x - y^2)(x - \sin y),$$
$$g(x, y) = (\cos y - x)(y - \cos x)$$

was designed to study the behavior of deflation when applied to a system which has several nearly equal, but distinct, zeros. This system has four zeros very close together in the first quadrant, and an infinite number of zeros spread else-

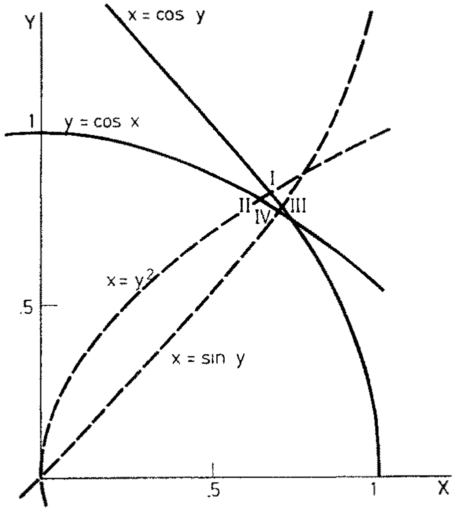

Fig. 2. Zero curves of the four-cluster system

where. The zeros of the four-cluster are $r_1 \simeq (0.68,\ 0.82)$, $r_2 \simeq (0.64,\ 0.80)$, $r_3 \simeq (0.71,\ 0.79)$ and $r_4 \simeq (0.69,\ 0.77)$. Experiments were run with eleven distinct starting guesses, the farthest away from any of the zeros being (0, 0). We noted the following results:

1. In no case were more than two members of the four-cluster found from the same starting guess, although other zeros not in the four-cluster were often found as well. Thus, although the deflation of one zero from the four-cluster was able to keep iterates away from that zero but not away from its very close neighbors, the weight of two deflated zeros from the four-cluster was sufficient to direct all iterates away from the general area of the four-cluster.

2. Different deflating techniques again produced convergence to different zeros of the four-cluster from the same starting guess. Thus with $x_0 = (0.9,\ 1)$, for example, Newton's Method converged to r_1 undeflated, but converged to r_2 using l_2 deflation, r_3 using l_∞ deflation and r_4 using gradient deflation.

3. Divergence to infinity occurred rarely, while divergence, in the form of nonconvergence in a prescribed number of iterations (50) occurred in about a quarter of the cases. In most cases, however, the algorithms continued to find zeros using the various deflation techniques until the maximum number of zeros specified by the programmer (4) were found.

Example 3. The Hyperbola-Circle (see Fig. 3)

The system

$$f(x, y) = xy - 1,$$
$$g(x, y) = x^2 + y^2 - 4,$$

has the four zeros $r_1 \simeq (0.517,\ 1.93)$, $r_2 \simeq (1.93,\ 0.517)$, $r_3 \simeq (-1.93,\ -0.517)$ and $r_4 \simeq (-0.517,\ -1.93)$. This system appeared at first to be unstable since, for

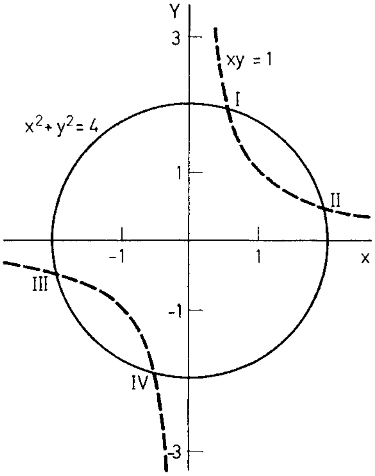

Fig. 3. Zero curves of the hyperbola-circle systeni

most starting guesses, Newton's Method diverged to infinity on the first attempt to find a zero. As it turned out, however, this system provided a major success for norm deflation (both types) in conjunction with Brown's method; for this combination found, from $x_0 = (0, 1)$, all four zeros of the system in quick succession. (Each of the two deflators took about 35 iterations to find all four zeros.) The l_2 deflator found the zeros in the order r_1, r_3, r_2, and r_4, while l_∞ deflation produced these roots in the order r_1, r_4, r_2, and r_3. The gradient deflator with Brown's method found only two zeros (r_1 and r_3) from this x_0 as did Newton's Method (r_1 and r_2 with each deflator).

Example 4. The 3×3 System

The zero-surface of

$$f(x, y, z) = x^2 + 2y^2 - 4$$

is an elliptic cylinder with elements parallel to the z-axis in 3-space; the zero surface of

$$g(x, y, z) = x^2 + y^2 + z - 8$$

is a circular paraboloid opening downward on the z-axis with vertex at (0, 0, 8); the zero surface of

$$h(x, y, z) = (x - 1)^2 + (2y - \sqrt{2})^2 + (z - 5)^2 - 4$$

is an ellipsoid having its center at the midpoint between the two zeros of the system which are given by $r_1 = (0, \sqrt{2}, 6)$ and $r_2 = (2, 0, 4)$. This problem was run using seven values of x_0 for Newton's Method and using two of the seven for Brown's method. The farthest starting guess from a solution for both methods was (1, 1, 1), which is about a distance of 5.1 (in Euclidean norm) from r_1. The other principle starting guess used was (1, 0.7, 5), which is approximately midway between the two zeros (of Euclidean distance $\simeq 1.58$ from each). In all cases which were run, both zeros were found in at most 19 iterations per zero, except for Newton's Method starting at (1, 1, 1) with l_2 deflation which failed to find the second zero in 200 iterations.

To summarize (a large number of numerical experiments), l_∞ norm deflation, besides being the easiest to compute, was the most stable of the three types of deflation studied.

6. Applications to Function Minimization

Given a real function of N real variables,

$$F(x_1, x_2, \ldots, x_N),$$

a technique often used to locate minima of F is to determine the zeros of the system of partial derivatives obtained by differentiating F with respect to each of its argument positions. This yields a nonlinear system of equations to be solved. The solutions of this system furnish candidates for the desired minimum points. The deflation techniques developed above may be used in avoiding such previously found minima.

Acknowledgement. The authors wish to thank Mr. H. A. Hall for his exacting and abundant efforts in conjunction with the numerical results presented in Section 5.

References

1. Brown, K. M.: Solution of simultaneous non-linear equations. Comm. Assoc. Comput. Mach. **10**, 728–729 (1967).
2. — A quadratically convergent Newton-like method based upon Gaussian elimination. SIAM J. Numer. Anal. **6**, 560–569 (1969).
3. Forsythe, G., Moler, C. B.: Computer solution of linear algebraic systems. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1967.
4. Wilkinson, J. H.: Rounding errors in algebraic processes. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1963.

Dr. Kenneth M. Brown*
Department of Computer Science
Cornell University
Ithaca, N.Y. 14850, USA

* On leave 1970–71 at:
Department of Computer Science
Yale University
New Haven, Conn. 06520, USA

Dr. William B. Gearhart
Department of Mathematics
The University of Utah
Salt Lake City, Utah 84112, USA