Implementing the initialization-dependence and the singularity difficulties in Newton's Method.

Theodoula N. Grapsa

June 2, 2009

Abstract

Newton's method is perhaps the most famous method for solving systems of nonlinear equations, because of its attractive advantage to converge quadratically if the corresponding Jacobian matrix is nonsingular and a good initial guess is available. In this report a mechanism of producing new initial points, named pivot-points, is introduced to reduce mainly the initial-dependence disadvantage of Newton's method.

1. Introduction

Solving systems of nonlinear equations is an important issue in the field of computational science and engineering. Nonlinear equations solving problems may arise directly in solving nonlinear systems or indirectly in optimization problems [10, ?]. In this paper we are concerned with the problem of solving systems of nonlinear equations

$$F(x) = \Theta^n = (0, 0, \ldots, 0), \tag{1}$$

where $F = (f_1, \ldots, f_n) : \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is a continuously differentiable mapping on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a solution $x^* \in \mathcal{D}$ of the system (1).

Several interesting methods have been proposed to contribute in this active area of research. The most famous of all is perhaps the so called Newton's method because of its quadratic convergence if some appropriate conditions are fulfilled. In particular, starting with an initial guess x^0 for the attainment of an approximation of the solution x^* of (1), the iteration of Newton's method are given by

$$x^{p+1} = x^p - J(x^p)^{-1} F(x^p), \quad p = 0, 1, \ldots. \tag{2}$$

If a good initial point is available and the Jacobian matrix is nonsingular Newton's method is quadratically convergence.

Despite of the good, in general, theoretical properties Newton's method has a number of disadvantages in practice. In particular:
–*Computation of Jacobian matrix.* The method may have high cost due to the computation of Jacobian matrix at each iteration. In such cases quasi-Newton methods are used to approximate the Jacobian matrix by other nonsingular one with less cost. The simplest modification of Newton method is the fixed Newton method, known as Modified Newton method [17]. Moreover, software packages for automatic differentiation have been produced and so this drawback is faced satisfactorily [17, 13]. But, quasi-Newton methods do not attain the quadratic convergence.
–*Solving of a linear system of equations.* Besides the computation of Jacobian matrix, a linear system must be solved at each iteration. In such cases inexact Newton methods are used to approximate the solution of such a system . As classical Newton's method, inexact Newton methods are also guaranteed to converge only if the initial approximation is closely to the solution, but without quadratic convergence [12, 14, 11] [M Pernice, R.D. Hornung—Newton Krylov-FAC methods for problems...].
–*Imprecise function values.* Newton's method remains problematic when the values of F can not be accurately achieved. Of course, this problem is common to all iterative procedures which directly depend on function evaluations. To overcome it, one may resort to generalized bisection methods [4, 15, 16, 25, 26, 27, 28] since they only make use of the algebraic sign of the function involved in the equations, but these methods do not generally attain a quadratic convergence. Moreover, we have proposed a class of methods, named Dimension Reducing

[6, 7], which work well with imprecise function values of F holding the quadratic convergence.
–*Singularity of Jacobian matrix.* If the Jacobian matrix is singular, Newton's method becomes problematic. In particular, when $J(x^*)$ is singular, it does not converge quadratically and, in general, it is not appropriate for approximations of x^* with a high accuracy. For this reason there are procedures which under some assumptions (such as rank $J(x^*) = n - 1$), can attain a highly accurate solution x^* by enlarging the system (1) to one which is at least $(2n + 1)$–dimensional[30, 31] (Kearfott). On the other hand, quadratic-convergent DR methods seems handling successfully this difficulty by reducing the system.
–*Initialization dependence.* Since the convergence behavior of Newton method is initialization dependent, it is reasonable Newton's method does not work well when the initial point is far away from the solution point. In this situation we have either a slow convergence or a divergence of the method. To be faced this situation, there are slow convergence methods, but initialization independent which are appropriate to used in early steps of the algorithm for finding starting points before a Newton-Type method is applied. Moreover, in practice, the initial problem, due to a bad initial guess, is faced by modifying the original problem so that a successful Newton's method may be applicable [].

Recapitulating from the above, the facts for DR method one may notice that it can accelerate the convergence even when a singularity exists or when a good initial guess is not available and moreover possible inaccurate function values of the function does not affect the overall behavior of DR method, since it is not directly depend of function values. But, although their attractive advantages, as we will see in the next section, DR method has been given in a rather complex form, in the sense that it works not on the trivial functions f_i of the original system but in a different equivalent set of functions v_i, whose analytic form is not generally available and only their values which are created during the process are known.

In this report, motivated from one hand by the attractive advantages of DR method that have to be kept and the rather inconvenient form of it that must be simplified and from the other hand by the need of reducing the problems in Newton's method due to initialization dependent and the singularity of Jacobian matrix, we propose a new Modified Newton method incorporating all the above, named Modified Newton Equivalent DR method (MNEDR). Actually, the new method, as we will see later, in some sense may be concerned as a simplified DR method although they work in different dimensions and in different functions. Moreover, the new method is a new Modified Newton method and the basic concept is a proper modification of Jacobian matrix, such that a proper rotation of Newton's tangent planes is becoming to accelerate the convergence reducing the number of iterations. Furthermore, the new Jacobian matrix is evaluated in appropriate points which lie on the solution surfaces C_i of functions f_i for $i = 1, 2, \ldots, n$ and have zero values. These points, in which we shall report in the sequel as pivot points, are the basic key idea to overcome the two assumption in order to keep quadratic convergence in Newton's method. This means that these pivot points, impinge to expand the domain convergence of Newton method,

succeeding fast convergence even from points far away from solution and besides impinge with success in the singularity problem of it. Mathematically, the pivot points arise from the solution of appropriate one dimensional equations. Their role is so great, due to such important advantages, that it is worth to pay to extract them. But, despite their acquisition cost, they impinge in a way, such that the method to be very fast convergent and at last the method to be generally lower cost, specifically in problematic cases.

Even though until now, there is any theoretical proof of why these pivot points have such well results for points far away from the solution, according to our knowledge we think that DR method and the new one may be considered as the first methods that does Newton's method, in some sense, initialization independent besides the improvement of its behavior in singularity problems.

Moreover, since the new proposed method may be faced as a simpler variant of DR method, it holds the main advantages of DR method but loses the advantage to work well with imprecise function values, since it works with the original functions f_i for $i = 1, 2, \ldots, n$ like Newton's method instead of DR ones we had proposed in DR method. This means that we rederive our DR method in a way to make it more evident and more practical, closer to Newton's form but with a little bit different advantages of DR's and it is at user disposal to decide the form he will use.

Also, the new proposed method like DR method may be easily parallelized since a componentwise technique is applied in each function f_i for $i = 1, 2, \ldots, n$ to extract the pivot points for Jacobian.

Furthermore, in this paper we propose algorithms for the DR method, as well as for the proposed one and besides we accent on the idea behind DR method, which had not given in [] and behind the new approach.

The rest of the paper is organized as follows. In section 2 the Newton and DR-method are very briefly given with their geometric interpretation , besides with the key idea behind DR method and the algorithm which had not given explicitly in [6, 7]. Furthermore, in the same section a mention in solving one dimensional equations is given, the essential tool of DR and new methods. In section 3 the new proposed Newton method is presented, with the presentation of the key idea behind it, its derivation and the new algorithm besides the proof of identification of DR and new one. At the end of this section the quadratic convergence proof of new algorithm and its geometric interpretation are given. In section 4 an analytic example and the geometric interpretation of it using Newton, DR and new Modified Newton methods are given. Furthermore, in this section by numerical applications it is shown that the solutions of systems of nonlinear equations could be found very efficiently by using the new proposed algorithm. The numerical results of the proposed Newton method and of the DR-method showed that at an early stage of the corresponding algorithm, even when the Jacobian is singular or the initial point is so far from the solution of the system of nonlinear equations, the new approximation goes quickly close to the solution of the system and in many applications from the first iteration.

2. The Pivot Points

The main drawback of Newton's method is the choice of an initial guess since it must be selected from the convergence region. That is, the convergence or the divergence of Newton's method depends immediately on a good initial point. In practical applications it is difficult such a point to be available.

Instead of guessing a single good initial point for all components f_i of the initial problem, we are seeking for different points for each f_i. In the following we propose some definitions and notions in order to generate those points.

Definition 2..1. *A point x^* is defined as a **solution point** of a system of nonlinear equations, if it satisfies*

$$f_i(x^*) = 0,$$

*for all $i \in \{1, 2, \ldots, n\}$. This feature of solution points is called **solution property**.*

Definition 2..2. *A point x^i_{pivot} has the **quasi-solution property** if it vanishes at least a function f_i, $i \in \{1, 2, \ldots, n\}$, that is $f_i(x) = 0$ and it is named **pivot point**.*

If a pivot-point x^i_{pivot} has the quasi-solution property and vanishes all the functions f_i, obviously, has the solution property too.

Definition 2..3. *The set of all pivot-points of a component f_i consists the solution surface c_{f_i} of f_i (zero-contour surface).*

Definition 2..4. *A pivot-point x_{pivot}, lying on the solution surface of a function is defined as candidate solution point.*

A candidate solution point x^i_{pivot} vanishes the corresponding components $f_i(x)$ for some $i \in \{1, 2, \ldots, n\}$ and has the quasi solution property. Hence, $f_i(x^i_{pivot}) = 0$ and generally for the rest components it is hold

$$f_j(x^i_{pivot}) \neq 0,$$

for $j \in \{1, 2, \ldots, i-1, i+1, \ldots, n\}$.

According to the above definitions, the solution surface of a function $f_i(x)$, for $i \in \{1, 2, \ldots, n\}$, may be considered as a set of all its candidate points, that is,

$$c_{f_i} = \{x \in D : f_i(x) = 0\} = U\left\{x^i_{pivot}\right\}.$$

In a theoretical point of view someone may choose any of the pivot-points from the corresponding f_i to substitute the initial guess. In this report we use the pivot-points that are lying in the solution surface and simultaneously in a line parallel to an axis of coordinate system, thus the selected pivot-points are their intersection.

Definition 2..5 (Pivot Set)**.** *We define as pivot set S_{pivot}, the set of all pivot point which correspond to a current point x^p. The number of elements of a pivot set is the number of intersection points of the line which is parallel to x_n-axis and the n functions $f_i(x)$. The maximum number of elements of a pivot set may be the number of functions.*

$$S_{pivot} = \left\{ x_{pivot}^{p,1}, x_{pivot}^{p,2}, \ldots, x_{pivot}^{p,n} \right\}. \tag{3}$$

3. Extraction of pivot set

Let

$$x^p = (x_1^p, x_2^p, ..., x_{n-1}^p, x_n^p) \tag{4}$$

be a current approximation of the solution $x^* = (x_1^*, x_2^*, ..., x_{n-1}^*, x_n^*)$ at an iteration $p = 0, 1, \ldots$ and let $y^p = (x_1^p, x_2^p, ..., x_{n-1}^p)$ the vector of the $n-1$ components of x^p.

Remark 3..1. *Without loss of generality we chose the n-th axis of the coordinate system to derive the pivot-points.*

From the chosen derivation process of pivot points it is obvious that the current point $x^p = (x_1^p, x_2^p, ..., x_{n-1}^p, x_n^p)$ and the corresponding pivot points $x_{pivot}^{p,i}$ have the same $n-1$ components $y^p = (x_1^p, x_2^p, ..., x_{n-1}^p)$ and differ only at the $n-th$ component, which is denoted by $x_n^{p,i}$. Hence,

$$x_{pivot}^{p,i} = (x_1^p, x_2^p, ..., x_{n-1}^p, x_n^{p,i}) = (y^p; x_n^{p,i}) \tag{5}$$

According to the Definition (2..2), a pivot point, as a candidate solution point of a function component $f_i(x)$, vanishes the corresponding function, and therefore the following relation is valid:

$$f_i(x_{pivot}^{p,i}) = 0, i = 1, \ldots, n. \tag{6}$$

At each current point x^p correspond n pivot points. One pivot point at each function $f_i(x), i = 1, 2, \ldots, n$.

As we earlier mentioned the development of new method demands on the pivot set S_{pivot} of n proper selected points $x_{pivot}^{p,i}$ at each iteration p. By the definition of pivot points they lie on the corresponding solution surface(**??**) and the relation (6) is valid.

Moreover, we selected them to be on the line which pass through the current point x_p in some iteration $p = 1, 2, \ldots, n$ and simultaneously it is parallel to some axis x_i, for some $i = 1, 2, \ldots, n$. In this paper we select the x_n axis. It is obvious that these points are the intersection of the solution surfaces c_{f_i} and the line $x = x_n^p$. From the way of construction of theses pivot points, they have the same $n-1$ components with the current point x_p and differ only at the nth component $x_n^{p,i}$, thus the relation (7) is valid.

Due to the way we define the pivot points, it is obvious that from mathematical point of view its remaining nth component, $x_n^{p,i}$, for each pivot point may be extracted by solving proper one-dimension equations, in the form

$$f_i(x_1^p, x_2^p, ..., x_{n-1}^p, .) = 0 \tag{7}$$

where the $n-1$ components $y_p = (x_1^p, x_2^p, \ldots, x_n^p)$ are fixed.

The solution of the one–dimensional equations $f_i(x_1^p, \ldots, x_{n-1}^p, \cdot) = 0$ may be done using any one of the well–known one–dimensional methods[20, 22, 24] to solve the above equations. Specifically here we use a simplified version of the bisection method, since it is sign-based-function and thus it requires only the algebraic signs of the function values to be correct. A further a discussion about its advantages can be found in[25, 26, 27, 28].

Remark If a function does not contain the component x_n then we substitute this function by another equivalent one. For example by adding in the function a function in which there exist the component x_n, or the pivot point of this function is taken the current point.

Notation 3..1. *Of course, relative procedures for obtaining x^* can be constructed by replacing x_n with any one of the components $x_1, \ldots, x_{n-1}$, for example x_{int}, and taking $y = (x_1, \ldots, x_{int-1}, x_{int+1}, \ldots, x_n)$.*

Bibliography

[1] Armijo L.(1996), "Minimization of functions having Lipschitz continuous first partial derivatives", Pacific J. Math. Vol. 16, pp. 1–3.

[2] Dennis J.E., Schnabel Jr. and R.B.(1983), *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice–Hall, Inc., Englewood Cliffs, New Jersey.

[3] Dieudonné J.(1969), *Foundations of modern analysis*, Academic Press, New York.

[4] Eiger A., Sikorski K. and Stenger F.(1984), "A bisection method for systems of nonlinear equations", ACM Trans. Math. Software, Vol. 10, pp. 367–377.

[5] Gorn A.(1954), "Maximal convergence intervals and a Gibb's–type phenomenon for Newton's approximation procedure", Ann. Math., Vol 59, pp. 463–476.

[6] Grapsa T.N. and Vrahatis M.N.(1989), "The implicit function theorem for solving systems of nonlinear equations in $\mathcal{R}^2$", Inter. J. Computer Math., Vol. 28, pp. 171–181.

[7] Grapsa T.N. and Vrahatis M.N.(1990), "A dimension–reducing method for solving systems of nonlinear equations in $\mathcal{R}^n$", Inter. J. Computer Math., Vol. 32, pp. 205–216.

[8] Grapsa T.N., Vrahatis M.N. and Bountis T.C.(1990), "Solving systems of nonlinear equations in $\mathcal{R}^n$ using a rotating hyperplane in $\mathcal{R}^{n+1}$", Inter. J. Computer Math., Vol. 35, 133–151.

[9] Grapsa T.N. and Vrahatis M.N.(1995), "A new dimension–reducing method for solving systems of nonlinear equations", Inter. J. Computer Math. Vol. 55, pp. 235–244.

[10] Grapsa T.N. and Vrahatis M.N.(1996), "A dimension–reducing method for unconstrained optimization", Journal of Computational and Applied Mathematics, Vol. 66, pp. 239–253.

[11] Grapsa T.N. and Vrahatis M.N.(1998), "inexact....",

[12] Dembo R.S., Eisenstat S.C. and Steihaug T.(1982), "Inexact Newton methods", SIAM J. Mumer. Anal., Vol. 19, pp. 400–408.

[13] Dixon L.C.(1991), "On the impact of automatic differentiation on the relative performance of parallel truncated Newton and variable metric algorithms", SIAM J. Optim., Vol 1, pp. 475–486.

[14] Eisenstat S.C. and Walker H.F.(1996), "Choosing the forcing terms in an inexact Newton method", SIAM J. Sci. Comput., Vol. 17, pp. 16–32.

[15] Kearfott R.B.(1979), "An efficient degree–computation method for a generalized method of bisec- tion", Numer. Math., Vol. 32, pp. 109–127.

[16] Kearfott R.B.(1987), "Some tests of generalized bisection", ACM Trans. Math. Software, Vol. 13, pp. 197–220.

[17] Krejic N. and Luzanin Z.(2001), "Newton-Like Method with Modification of the Right-Hand-Side Vector", Mathematics of Computation, Vol. 71, No 237, pp. 237–250.

[18] Kupferschmid M. and Ecker J.G.(1987), "A note on solution of nonlinear programming problems with imprecise function and gradient values", Math. Program. Study, Vol 31, pp. 129–138.

[19] Moré B.J., Garbow B.S. and Hillstrom K.E.(1981), "Testing unconstrained optimization", ACM Trans. Math. Software, Vol. 7, pp. 17–41.

[20] Ortega J.M. and Rheinbolt W.C.(1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

[21] Press W.H., Teukolsky S.A., Vetterling W.T. and Flannery B.P.(1992), *Numerical Recipes, The Art of Scientific Computing, 2nd ed.*, Cambridge University Press, New York.

[22] Rheinboldt W.C.(1974), *Methods for solving systems of equations*, SIAM, Philadelphia.

[23] Sikorski K.(1982), "Bisection is optimal", Numer. Math., Vol. 40, pp. 111–117.

[24] Traub J.F.(1964), *Iterative methods for the solution of equations*, Prentice–Hall., Inc., Englewood Cliffs, NJ.

[25] Vrahatis M.N.(1988), "Solving systems of nonlinear equations using the nonzero value of the topological degree", ACM Trans. Math. Software, Vol. 14, pp. 312–329.

[26] Vrahatis M.N.(1988), "CHABIS: A mathematical software package for locating and evaluating roots of systems of nonlinear equations", ACM Trans. Math. Software, Vol. 14, pp. 330–336.

[27] Vrahatis M.N.(1989), "A short proof and a generalization of Miranda's existence theorem", Proc. Amer. Math. Soc., Vol. 107, pp. 701–703.

[28] Vrahatis M.N. and Iordanidis K.I.(1986), "A rapid generalized method of bisection for solving systems of non–linear equations", Numer. Math., Vol. 49, pp. 123–138.

[29] Vrahatis M.N., G.S. Androulakis and G.E. Manoussakis (1996), "A new unconstrained optimization method for imprecise function and gradient values", J. Math. Anal. Appl., Vol. 197, No 2, pp. 586-607

[30] Yamamoto N.(1984), "Regularization of solutions of nonlinear equations with singular Jacobian matrices", J. Infor. Processing, Vol. 7, pp. 16–21.

[31] Weber H. and Werner W.(1981), "On the accurate determination of nonisolated solutions of nonlinear equations", Computing, Vol. 26, pp. 315–326.