

FINDING GOOD STARTING POINTS FOR SOLVING EQUATIONS BY  
NEWTON'S METHOD

IOANNIS K. ARGYROS\*

**Abstract.** We study the problem of finding good starting points for the semilocal convergence of Newton's method to a locally unique solution of an operator equation in a Banach space setting. Using a weakened version of the Newton–Kantorovich theorem we show that the procedure suggested by Kung [6] is improved in the sense that the number of Newton-steps required to compute a good starting point can be significantly reduced (under the same computational cost required in the Newton–Kantorovich theorem [3], [5]).

**MSC 2000.** 65H10, 65G99, 47H17, 49M15.

**Keywords.** Newton's method, Banach space, Newton–Kantorovich theorem/hypothesis, Fréchet-derivative, Lipschitz, center–Lipschitz conditions, good starting points for Newton's method.

## 1. INTRODUCTION

In this study we are concerned with the problem of approximating a locally unique solution  $x^*$  of equation

$$(1.1) \quad F(x) = 0,$$

where  $F$  is a Fréchet-differentiable operator defined on an open convex subset  $D$  of a Banach space  $X$  with values in a Banach space  $Y$ .

A large number of problems in applied mathematics and also in engineering are solved by finding the solutions of certain equations. For example, dynamic systems are mathematically modeled by difference or differential equations, and their solutions usually represent the states of the systems. For the sake of simplicity, assume that a time-invariant system is driven by the equation  $\dot{x} = Q(x)$  (for some suitable operator  $Q$ ), where  $x$  is the state. Then the equilibrium states are determined by solving equation (1.1). Similar equations are used in the case of discrete systems. The unknowns of engineering equations can be functions (difference, differential, and integral equations), vectors (systems of linear or nonlinear algebraic equations), or real or complex numbers (single algebraic equations with single unknowns). Except in special cases, the most commonly used solution methods are iterative - when starting from

---

\* Cameron University, Department of Mathematics Sciences, Lawton, OK 73505, USA, e-mail: [iargyros@cameron.edu](mailto:iargyros@cameron.edu).

one or several initial approximations a sequence is constructed that converges to a solution of the equation. Iteration methods are also applied for solving optimization problems. In such cases, the iteration sequences converge to an optimal solution of the problem at hand. Since all of these methods have the same recursive structure, they can be introduced and discussed in a general framework.

The most popular method for generating a sequence approximation  $x^*$  is undoubtedly Newton's method:

$$(1.2) \quad x_{n+1} = x_n - F'(x_n)^{-1}F(x_n) \quad (n \geq 0) \quad (x_0 \in D).$$

Here  $F'(x)$  is the Fréchet derivative of  $F$  which is a bounded linear operator from  $X$  into  $Y$ . A survey of local and semilocal convergence results for Newton's method (1.2) can be found in [1], [3], [5] and the references there.

In particular the famous Newton–Kantorovich theorem guarantees the quadratic convergence of method (1.2) if the initial guess  $x_0$  is “close enough” to the solution  $x^*$  [5] (see (2.11)).

However we recently showed that the Newton–Kantorovich hypothesis (2.11) can always be replaced by the weaker (2.5) (under the same computational cost) [1, p. 387, Case 3, for  $\delta = \delta_0$ ]. In particular using the algorithm proposed by H.T. Kung [6] (see also [7]) we show that the number of steps required to compute a good starting point  $x_0$  (to be precised later) can be significantly reduced.

This observation is very important in computational mathematics [3], [5], [7].

## 2. FINDING GOOD STARTING POINTS FOR NEWTON'S METHOD (1.2)

We recently showed the following semilocal convergence theorem for Newton's method (1.2) which essentially states the following [1]:

If:

$$(2.1) \quad F'(x_0)^{-1} \text{ exists, } \|F'(x_0)^{-1}\| \leq \beta_0,$$

$$(2.2) \quad \|F'(x_0)^{-1}F(x_0)\| \leq \xi_0,$$

$$(2.3) \quad \|F'(x) - F'(x_0)\| \leq K_0 \|x - x_0\|,$$

$$(2.4) \quad \|F'(x) - F'(y)\| \leq K \|x - y\|,$$

for all  $x, y \in \bar{U}(x_0, r) = \{x \in X \mid \|x - x_0\| \leq r\}$ ,

$$(2.5) \quad h_0 = \beta_0 L \xi_0 < \frac{1}{2},$$

where,

$$(2.6) \quad L = \frac{1}{8} (K + 4 K_0 + \sqrt{K^2 + 8 K_0 K}),$$

$$(2.7) \quad a \xi_0 \leq r, \quad a = \frac{2}{2-b}, \quad b = \frac{1}{2} \left[ -\frac{K}{K_0} + \sqrt{\left(\frac{K}{K_0}\right)^2 + 8 \frac{K}{K_0}} \right],$$

and

$$(2.8) \quad \bar{U}(x_0, r) \subseteq D,$$

then sequence  $\{x_n\}$  ( $n \geq 0$ ) generated by Newton's method (1.2) is well defined, remains in  $\bar{U}(x_0, r)$  for all  $n \geq 0$  and converges quadratically to a unique solution  $x^* \in U(x_0, r)$  of equation  $F(x) = 0$ . Moreover we have

$$(2.9) \quad \|x_0 - x^*\| \leq a \xi_0.$$

REMARK 1. In general

$$(2.10) \quad K_0 \leq K$$

holds. If equality holds in (2.10) then the result stated above reduces to the famous Newton–Kantorovich theorem and (2.5) to the Newton–Kantorovich hypothesis (2.11) [5]. If strict inequality holds in (2.10) then (2.5) is weaker than the Newton–Kantorovich hypothesis

$$(2.11) \quad \bar{h}_0 = \beta_0 K \xi_0 < \frac{1}{2}.$$

Moreover the error bounds on the distances  $\|x_{n+1} - x_n\|$ ,  $\|x_n - x^*\|$  ( $n \geq 0$ ) are finer and the information on the location of the solution more precise [1]. In [1] examples were also given to show that (2.11) is violated but (2.5) holds. Furthermore as the following example demonstrates  $\frac{K}{K_0}$  can be arbitrarily large:  $\square$

EXAMPLE 1. Let  $X = Y = \mathbb{R}$ ,  $x_0 = 0$  and define function  $F$  on  $\mathbb{R}$  by

$$(2.12) \quad F(x) = c_0 x + c_1 + c_2 \sin e^{c_3 x}$$

where  $c_i$ ,  $i = 0, 1, 2, 3$  are given parameters. Using (2.3), (2.4) and (2.12) we can easily see that for  $c_3$  large and  $c_2$  sufficiently small  $\frac{K}{K_0}$  can be arbitrarily large.

That is (2.5) may hold but not (2.11).

Note also that the computational cost of obtaining  $(K_0, K)$  is the same as the one for  $K$  since in practice evaluating  $K$  requires finding  $K_0$ .

Hence all results using (2.11) instead of (2.5) can now be challenged to obtain more information. That is exactly what we are doing here. In particular motivated by the elegant work of H.T. Kung [6] on good starting points for Newton's method we show how to improve on these results if we use our theorem stated above instead of the Newton–Kantorovich theorem.  $\square$

DEFINITION 1. We say  $x_0$  is a good starting point for approximating  $x^*$  by Newton's method or a good starting point for short if conditions (2.1)–(2.8) hold.

Note that the existence of a good starting point implies the existence of a solution  $x^*$  of equation  $F(x) = 0$  in  $U(x_0, a \xi_0)$ .

We provide the following theorem/Algorithm which improves the corresponding ones given in [6, Theorem 4.1] to obtain good starting points.

**THEOREM 1.** *Let  $F: D \subseteq X \rightarrow Y$  be a Fréchet-differentiable operator. If  $F'$  satisfies center-Lipschitz, Lipschitz conditions (2.3), (2.4) respectively on  $U(x_0, 2r)$*

$$(2.13) \quad \begin{aligned} & \|F(x_0)\| \leq \eta_0, \\ & \|F'(x)^{-1}\| \leq \beta \quad \text{for all } x \in U(x_0, 2r), \\ & U(x_0, 2r) \subseteq D, \end{aligned}$$

and

$$(2.14) \quad \beta \eta_0 < \frac{r}{2},$$

then there exists a solution  $x^*$  of equation  $F(x) = 0$  in  $U(x_0, 2r)$ .

*Proof.* Simply use  $L$  instead of  $K$  in the proof of Theorem 4.1 in [6, p. 11] including the algorithm there which is essentially repeated here with some modifications:

**Algorithm A:** The goal of this algorithm is to produce a good starting point (under the conditions of Theorem 1) for approximating  $x^*$ .

- (1) Set  $h_0 \leftarrow \beta^2 L \eta_0$  and  $i \leftarrow 0$ . Choose any number  $\delta$  in  $(0, \frac{1}{2})$ .
- (2) If  $h_i < \frac{1}{2}$ ,  $x_i$  is a good starting point for approximating  $x^*$  and Algorithm A terminates.
- (3) Set  $\lambda_i \leftarrow (\frac{1}{2} - \delta)/h_i$ , and

$$F_i(x) \leftarrow [F(x) - F(x_i)] + \lambda_i F(x_i).$$

- (4) (It is shown in the proof that  $x_i$  is a good starting point for approximating a zero, denoted by  $x_{i+1}$ , of  $F_i$ .) Apply Newton's method to  $F_i$ , starting from  $x_i$ , to find  $x_{i+1}$ .
- (5) (Assume that the exact  $x_{i+1}$  is found.) Set  $\eta_{i+1} \leftarrow \|F(x_{i+1})\|$  and

$$h_{i+1} \leftarrow \beta^2 K \eta_{i+1}.$$

- (6) Set  $i \leftarrow i + 1$ , and return back to Step 2.

That completes the proof of Theorem 1. □

**REMARK 2.** As already noted in [6] Theorem 1 is trivial for the scalar case ( $f: \mathbb{R} \rightarrow \mathbb{R}$ ), since the mean value theorem can be used. Some of the assumptions of Theorem 1 can be weakened. Avila for example in [4, Theorem 4.3] instead of (2.14) used a more complicated condition involving  $\beta$ ,  $K$ , and  $\eta_0$ . However the idea of his algorithm is basically different from Algorithm A. Note also that if  $K_0 = K$  then our Theorem 1 reduces to Theorem 4.1 in [6]. We now modify Algorithm A to make it work in Banach spaces without necessarily assuming that the exact zero of  $x_{i+1}$  of  $F_i$  can be found using Newton's method (1.2). □

**THEOREM 2.** *Under the hypotheses of Theorem 1 a good starting point for approximating solution  $x^*$  of equation  $F(x) = 0$  can be obtained in  $N(\delta, K_0, K)$  Newton steps,  $\delta$  is any number in  $(0, \frac{1}{2})$ ,*

$$(2.15) \quad N(\delta, K_0, K) = \begin{cases} 0, & \text{if } h_0 = \beta^2 L \eta_0 \leq \frac{1}{2} - \delta, \\ I(\delta, K_0, K) \cdot J(\delta, K_0, K), & \text{otherwise} \end{cases}$$

where,  $I(\delta, K_0, K)$  is the smallest integer  $i$  such that:

$$(2.16) \quad \left[ 1 - \frac{\frac{1}{2} - \delta}{h_0} \right]^i \leq \frac{[\frac{1}{2} - \delta]}{h_0},$$

and  $J(\delta, K_0, K)$  is the smallest integer  $j$  such that:

$$(2.17) \quad 2^{\frac{1}{j-1}} (1 - 2\delta)^{2^{j-1}} (a + \beta \eta_0) \leq r - 2\beta \eta_0,$$

$$(2.18) \quad \frac{1}{2^j} (1 - 2\delta)^{2^{j-1}} (a + \beta \eta_0) \leq a,$$

where,

$$(2.19) \quad a = \min \left( \frac{r}{2} - \beta \eta_0, \frac{\delta}{2\beta L} \right).$$

*Proof.* Simply use  $L$  instead of  $K$  in the proof of Theorem 4.2 in [6, p. 16], and the following algorithm.

**Algorithm B.**

- (1) Set  $h_0 \leftarrow \beta^2 L \eta_0$ ,  $\bar{x}_0 \leftarrow x_0$  and  $i \leftarrow 0$ . Choose any number  $\delta$  in  $(0, \frac{1}{2})$ .
- (2) If  $h_i \leq \frac{1}{2} - \delta$ ,  $\bar{x}_i$  is a good starting point for approximating  $x^*$  and Algorithm B terminates.
- (3) Set  $\lambda_i \leftarrow \frac{(\frac{1}{2} - \delta)}{h_i}$ ,

$$F_i(x) \leftarrow \frac{F(x) - \eta_i F(x_0)}{\eta_0} + \frac{\lambda_i \eta_i F(x_0)}{\eta_0},$$

and

$$\eta_{i+1} \leftarrow (1 - \lambda_i) \eta_i.$$

- (4) Apply Newton's method to  $F_i$ , starting from  $\bar{x}_i$ , to find an approximation  $\bar{x}_{i+1}$  to a zero  $x_{i+1}$  of  $F_i$  such that

$$\|\bar{x}_{i+1} - x_{i+1}\| \leq r - 2\beta \eta_0,$$

and

$$\|F'_i(\bar{x}_{i+1})^{-1} F_i(\bar{x}_{i+1})\| \leq \min \left( \frac{r}{2} - \beta \eta_{i+1}, \frac{\delta}{2\beta L} \right).$$

- (5) Set  $h_{i+1} \leftarrow \beta^2 L \eta_{i+1}$ .
- (6) Set  $i \leftarrow i + 1$  and return to Step 2.

That completes the proof of Theorem 2.  $\square$

**REMARK 3.** As noted in [6]  $\delta$  should not be chosen to minimize the complexity of Algorithm B. Instead,  $\delta$  should be chosen to minimize the complexity of algorithm:

- (1) Search Phase: Perform Algorithm B.
- (2) Iteration Phase: Perform Newton's method starting from the point obtained by Algorithm B.

An upper bound on the complexity of the iteration phase is the time needed to carry out  $T(\delta, K_0, K, \varepsilon)$  Newton steps, where  $T(\delta, K_0, K, \varepsilon)$  is the smallest integer  $k$  such that

$$(2.20) \quad \frac{1}{2^{k-1}}(1 - 2\delta)^{2^k-1}(a + \beta \eta_0) \leq \varepsilon.$$

Note also that if  $K_0 = K$  our Theorem 2 reduces to Theorem 4.2 in [6, p. 15].  $\square$

Hence we showed the following result:

**THEOREM 3.** *Under the hypotheses of Theorem 2 the time needed to find a solution  $x^*$  of equation  $F(x) = 0$  inside a ball of radius  $\varepsilon$  is bounded above by the time needed to carry out  $R(\delta, K_0, K, \varepsilon)$  Newton steps, where*

$$(2.21) \quad R(\varepsilon, K_0, K) = \min_{0 < \delta < \frac{1}{2}} [N(\delta, K_0, K, \varepsilon) + T(\delta, K_0, K, \varepsilon)],$$

where  $N(\delta, K_0, K, \varepsilon)$  and  $T(\delta, K_0, K, \varepsilon)$  are given by (2.15) and (2.20) respectively.

**REMARK 4.** If  $K_0 = K$  Theorem 3 reduces to Theorem 4.3 in [6, p. 20]. In order for us to compare our results with the corresponding ones in [6] we computed the values of  $R(\varepsilon, K_0, K)$  for  $F$  satisfying the conditions of Theorem 4.3 in [6] and Theorem 3 above with

$$(2.22) \quad \beta \eta_0 \leq .4r,$$

and

$$(2.23) \quad 1 \leq h_0 = \beta^2 L \eta_0 \leq 10,$$

and for  $\varepsilon$  equal to  $10^{-i}r$ ,  $1 \leq i \leq 10$ .

The following table gives the results for  $\varepsilon = 10^{-6}r$ . Note that by  $I$  we mean  $I(\delta_0, K, K)$ ,  $I_{\alpha K}$  we mean  $I(\delta_0, K_0, K)$  with  $K_0 = \alpha K$ ,  $\alpha \in [0, 1]$ . Similarly for  $J$ ,  $N$  and  $T$ . Note also that the columns  $J_{.9K}$ ,  $J_{.5K}$ ,  $J_{0K}$  as identical to  $J$  have been omitted. The columns  $I$ ,  $J$ ,  $N$ ,  $T$ ,  $R$  were given in [6], where they did not use  $K_0$  but  $K$  in (2.3).

The table was produced using (2.15)–(2.23):

$h_0$	$\delta_0$	I	J	N	T	R	I. $_{.9K}$	N. $_{.9K}$	R. $_{.9K}$	I. $_{.5K}$	N. $_{.5K}$	R. $_{.5K}$	I $_{0K}$	N $_{0K}$	R $_{0K}$
1	.165	3	2	6	5	11	3	6	11	2	4	9	1	2	8
2	.103	8	3	24	6	30	7	21	27	5	15	21	2	6	12
3	.118	16	3	48	6	54	14	42	48	10	30	36	5	15	21
4	.129	25	3	75	6	81	23	69	75	16	48	54	9	27	33
5	.137	35	3	105	6	111	33	99	105	30	90	96	13	39	45
6	.144	47	3	141	5	146	44	132	137	31	93	98	17	51	56
7	.149	59	3	177	5	182	55	165	170	40	120	125	22	66	71
8	.154	72	3	216	5	221	67	201	206	41	123	128	28	84	89
9	.159	85	3	255	5	260	80	240	245	58	174	179	33	99	104
10	.163	99	3	295	5	302	93	279	284	68	204	209	39	117	122

□

REMARK 5. It follows from the table that our results (see columns  $I_{.qK}$  and after) significantly improve the corresponding ones in [6] and under the same computational cost. Suppose for example that  $h_0 = 9$ ,  $\delta = .159$ . Kung found that the search phase can be done in 255 Newton steps and the iteration phase in 5 Newton steps. That is a root can be located inside a ball of radius  $10^{-6}r$  using 260 Newton steps. However for  $K_0 = .9K$ ,  $K_0 = .5K$  and  $K_0 = 0$  the corresponding Newton steps are 245, 179 and 104 respectively which constitute a significant improvement. At the end of his paper Kung asked whether the number of Newton steps used by this procedure is close to the minimum. It is now clear from our approach that the answer is no (in general).

Finally Kung proposed the open question: Suppose that the conditions of the Newton–Kantorovich theorem hold: Is Newton’s method optimal or close to optimal, in terms of the numbers of function and derivative equations required to approximate the solution  $x^*$  of equation  $F(x) = 0$  to within a given tolerance  $\varepsilon$ ?

Clearly according to our approach the answer is no (see also Remark 1). □

REMARK 6. The results obtained here further improve the corresponding ones given by us in [2]. There, we used

$$\bar{L} = \frac{K_0 + K}{2},$$

instead of  $L$  in (2.6).

However, we have

$$\bar{L} \leq L \quad \text{for all } K_0 \geq 0 \quad \text{and } K \geq 0.$$

If  $L = \bar{L}$ , then the results obtained here coincide with the corresponding ones in [1] (which in turn improved the ones in [4], [6], [7]). Otherwise (i.e. if  $\bar{L} < L$ ), then the results in this study improve our results in [2]. □

**REFERENCES**

1. ARGYROS, I.K., *A unifying local-semilocal convergence analysis and applications for two-point Newton-like methods in Banach space*, J. Math. Anal. Applic., **298**, 2, pp. 374–397, 2004.
2. ARGYROS, I.K., *An improved approach of obtaining good starting points for solving equations by Newton's method*, Adv. Nonlinear Var. Ineq., **8**, pp. 133–142, 2005.
3. ARGYROS, I.K., *Computational theory of iterative methods*, Studies in Computational Mathematics, **15**, Elsevier, 2007, New York, U.S.A.
4. AVILA, J., *Continuation method for nonlinear equations*, Technical Report TR-142, Computer Science Center, University of Maryland, January, 1971.
5. KANTOROVICH, L.V. and AKILOV, G.P. *Functional Analysis in Normed Spaces*, Pergamon Press, Oxford, 1982.
6. KUNG, H.T., *The complexity of obtaining starting points for solving operator equations by Newton's method*, Technical Report NR 044-422, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA, 15213, October, 1975.
7. TRAUB, J.F. (editor), *Analytic Computational Complexity*, Academic Press, 1975.

Received by the editors: May 12, 2008.