The Secant-Newton Map is Optimal Among Contracting Quadratic Maps for Square Root Computation *

Mădălina Eraşcu†
Research Institute for Symbolic Computation, Johannes Kepler University, A-4040, Linz, Austria
merascu@risc.jku.at

Hoon Hong
Department of Mathematics, North Carolina State University, Box 8205, Raleigh NC 27695, USA
hong@ncsu.edu

Abstract

Consider the problem: given a real number x and an error bound ε, find an interval such that it contains $\sqrt{x}$ and its width is less than ε. One way to solve the problem is to start with an initial interval and repeatedly to update it by applying an interval refinement map on it until it becomes narrow enough. In this paper, we prove that the well known Secant-Newton map is the *optimal* among a certain family of natural generalizations.

Keywords: square root, interval, secant, Newton, contracting
AMS subject classifications: 65G20, 65G30

1 Introduction

Computing the square root of a given real number is a fundamental operation. Naturally, various numerical methods have been developed [4, 9, 11, 6, 7, 5, 3, 1, 2, 10, 8]. In this paper, we consider an interval version of the problem [7, 1, 8]: given a real number x and an error bound ε, find an interval such that it contains $\sqrt{x}$ and its width is less than ε. One way to solve the problem starts with an initial interval and repeatedly updates it by applying a *refinement* map, say R, on it until it becomes narrow enough (see below).

*Submitted: August 31, 2012; Revised: March 11, 2013; Accepted: May 26, 2013; Posted: May 26, 2013.
†Recipient of a DOC-fFORTE-fellowship of the Austrian Academy of Sciences

in: $x > 0,\ \ \varepsilon > 0$
out: I, interval such that $\sqrt{x} \in I$ and $\mathsf{width}(I) \leq \varepsilon$

$I \leftarrow [\min(1, x), \max(1, x)]$
while $\mathsf{width}(I) > \varepsilon$
    $I \leftarrow R(I, x)$
return I

A well known refinement map R, tailored for square root, is obtained by combining the secant map and the Newton map where the secant/Newton map is used for determining the lower/upper bound of the refined interval, that is,

$$R: \ [L, U], x \ \mapsto \ \left[L + \frac{x - L^2}{L + U}, U + \frac{x - U^2}{2U}\right]$$

which can be easily derived from Figure 1.[1] [2] A question naturally arises. *Is there any*

Figure 1: Derivation of Secant-Newton map

refinement map which is better than Secant-Newton? In order to answer the question rigorously, one first needs to fix a search space, that is, a family of maps in which we search for a better map. In this paper, we will consider the family of all the "natural generalization" of Secant-Newton map. The above picture shows that Secant-Newton map is contracting, that is, $L \leq L' \leq \sqrt{x} \leq U' \leq U$. Furthermore, it "scales properly", that is, if we multiply $\sqrt{x},\ L$ and U by a number, say s, then L' and U' are also multiplied by s. This is due to the fact that the numerators are quadratic forms in $\sqrt{x},\ L$ and U and the denominators are linear forms. These observations suggest the following choice of a search space: the family of all the maps with the form

$$R: \ [L, U], x \mapsto [L', U']$$
$$L' = L + \frac{x + p_0 L^2 + p_1 LU + p_2 U^2}{p_3 L + p_4 U}$$
$$U' = U + \frac{x + q_0 U^2 + q_1 UL + q_2 L^2}{q_3 U + q_4 L}$$

such that

$$L \leq L' \leq \sqrt{x} \leq U' \leq U$$

[1] An anonymous referee made an interesting observation that the Secant-Newton map can be also viewed as an instance of the interval Newton map with slope:

$$[L, U], x \ \mapsto \ m - \frac{m^2 - x}{m + [L, U]}$$

where $m \in [L, U]$. If we choose $m = U$ then it is identical to the Secant-Newton map.

[2] It is important to note that there are faster non-interval algorithms for computing square roots [6, 5, 3, 2]. They are based on static error analysis, auto-corrective behavior of Newton map, etc. However, in this paper, we restrict our investigation to interval methods because the current work is carried out as a preliminary study, in the hope of identifying conceptual and technical tools for finding an optimal method for solving polynomial equations. Interval based methods have the benefit of providing a uniform paradigm for such larger class of problems.

which we will call *contracting quadratic* maps. By choosing the values for the parameters $p = (p_0, \ldots, p_4)$ and $q = (q_0, \ldots, q_4)$, we get each member of the family. For instance, Secant-Newton map can be obtained by setting $p = (-1, 0, 0, 1, 1)$ and $q = (-1, 0, 0, 2, 0)$.

The main contribution of this paper is the finding that Secant-Newton map is the *optimal* among all the contracting quadratic maps. By optimal, we mean that the output interval of Secant-Newton map is always proper subset of that of all the other contracting quadratic maps, as long as $\sqrt{x}$ resides in the interior of the input interval.

The paper is structured as follows. In Section 2, we precisely state the main claim of the paper. In Section 3, we prove the main claim.

2 Main Result

In this section, we will make a precise statement of the main result informally described in the previous section. For this, we recall a few notations and notions.

We say that a map

$$R : [L, U], x \;\mapsto\; [L', U']$$

is a quadratic map *if it has the following form*[3]

$$L' = L + \frac{x + p_0 L^2 + p_1 LU + p_2 U^2}{p_3 L + p_4 U}$$
$$U' = U + \frac{x + q_0 U^2 + q_1 UL + q_2 L^2}{q_3 U + q_4 L}$$

We will denote such a map by $R_{p,q}$.

Definition 2 (Secant-Newton map)**.** *The Secant-Newton map is the quadratic map* R_{p^*,q^*} *where* $p^* = (-1, 0, 0, 1, 1)$ *and* $q^* = (-1, 0, 0, 2, 0)$*, namely*

$$R_{p^*,q^*} : [L, U], x \;\mapsto\; [L^*, U^*]$$

where

$$L^* = L + \frac{x - L^2}{L + U}$$
$$U^* = U + \frac{x - U^2}{2U}$$

Definition 3 (Contracting quadratic map)**.** *We say that a map*

$$R : [L, U], x \;\mapsto\; [L', U']$$

is a contracting quadratic map *if it is a quadratic map and*

$$\underset{L,U,x}{\forall} \;\; 0 < L \leq \sqrt{x} \leq U \implies L \leq L' \leq \sqrt{x} \leq U' \leq U \tag{1}$$

Now we are ready to state the main result of the paper.

[3] A careful reader would be concerned about the possibility of the denominators becoming 0, making the expressions undefined. Fortunately it will turn out that these cases will be naturally eliminated in the subsequent discussions.

Theorem 1 (Main Result)**.** *Let $R_{p,q}$ be a contracting quadratic map which is not R_{p^*,q^*} (Secant-Newton). Then we have*

$$(a) \quad \underset{L,U,x}{\forall} \quad 0 < L \leq \sqrt{x} \leq U \quad \Longrightarrow \quad R_{p^*,q^*}([L,U],x) \quad \subseteq \quad R_{p,q}([L,U],x)$$

$$(b) \quad \underset{L,U,x}{\forall} \quad 0 < L < \sqrt{x} < U \quad \Longrightarrow \quad R_{p^*,q^*}([L,U],x) \quad \subsetneq \quad R_{p,q}([L,U],x)$$

Remark 1. *It is important to pay a careful attention to a subtle difference between the two claims (a) and (b). In the first claim, $\sqrt{x}$ is allowed to lie on the boundary of the input interval, namely $\sqrt{x} = L$ or $\sqrt{x} = U$. In the second claim, $\sqrt{x}$ is required to lie in the interior of the input interval.*

Remark 2. *The first claim states that Secant-Newton map is never worse than any other contracting quadratic map as along as $\sqrt{x}$ resides in the input interval. The second claim states that Secant-Newton map is always better than all the other contracting quadratic maps as long as $\sqrt{x}$ resides in the interior of the input interval.*

3 Proof

In this section, we will prove the main result (Theorem 1). For the sake of easy readability, the proof will be divided into several lemmas, which are interesting on their own. The main theorem follows immediately from the last two lemmas (Lemmas 6 and 7).

Lemma 2. *Let $R_{p,q}$ be a contracting quadratic map. Then we have*

$$\begin{array}{lll} p_0 = -1 & p_1 = 0 & p_2 = 0 \\ q_0 = -1 & q_1 = 0 & q_2 = 0 \end{array}$$

Proof. Let $R_{p,q}$ be a contracting quadratic map. Then p, q satisfy the condition (1). The proof essentially consist of instantiating the condition (1) on $x = L^2$ and $x = U^2$.

By instantiating the condition (1) with $x = L^2$ and recalling the definition of L', we have

$$\underset{L,U}{\forall} \quad 0 < L \leq U \implies L + \frac{L^2 + p_0 L^2 + p_1 LU + p_2 U^2}{p_3 L + p_4 U} = L$$

By simplifying, removing the denominator and collecting, we have

$$\underset{L,U}{\forall} \quad (L,U) \in D \implies g\,(L,U) = 0$$

where

$$D = \{(L,U) : 0 < L \leq U\}$$
$$g\,(L,U) = (1 + p_0)L^2 + p_1 LU + p_2 U^2$$

Since the bivariate polynomial g is zero over the 2-dim real domain D, it must be identically zero. Thus its coefficients $1 + p_0$, p_1 and p_2 must be all zero.

By instantiating the condition (1) with $x = U^2$ and recalling the definition of U', we have

$$\underset{L,U}{\forall} \quad 0 < L \leq U \implies U + \frac{U^2 + q_0 U^2 + q_1 UL + q_2 L^2}{q_3 U + q_4 L} = U$$

By simplifying, removing the denominator and collecting, we have

$$\underset{L,U}{\forall} \ (L,U) \in D \implies h\,(L,U) = 0$$

where

$$D = \{(L,U) : 0 < L \leq U\}$$
$$h\,(L,U) = (1+q_0)\,U^2 + q_1 UL + q_2 L^2$$

Since the bivariate polynomial h is zero over the 2-dim real domain D, it must be identically zero. Thus its coefficients $1+q_0,\ q_1$ and q_2 must be all zero. □

Lemma 3. *Let $R_{p,q}$ be a contracting quadratic map. Then we have*

$$L' = L + \frac{x - L^2}{p_3 L + p_4 U}$$
$$U' = U + \frac{x - U^2}{q_3 U + q_4 L}$$

Proof. Let $R_{p,q}$ be a contracting quadratic map. From Lemma 2, we have

$$p_0 = -1 \qquad p_1 = 0 \qquad p_2 = 0$$
$$q_0 = -1 \qquad q_1 = 0 \qquad q_2 = 0$$

Recalling the definition of L' and U', we have

$$L' = L + \frac{x - L^2}{p_3 L + p_4 U}$$
$$U' = U + \frac{x - U^2}{q_3 U + q_4 L}$$

□

The following lemma will be used to simplify the proof of Lemma 5.

Lemma 4. *If*

$$\underset{X,Y,Z}{\forall} \ 0 < X < Y < Z \implies aX + bY + cZ \geq 0$$

then

$$a+b+c \geq 0 \qquad b+c \geq 0 \qquad c \geq 0$$

Proof. Assume

$$\underset{X,Y,Z}{\forall} \ 0 < X < Y < Z \implies aX + bY + cZ \geq 0$$

Let $x = X,\ y = Y - X$ and $z = Z - Y$. Then we can rewrite the above as

$$\underset{x,y,z}{\forall} \ x,y,z > 0 \implies ax + b\,(x+y) + c\,(x+y+z) \geq 0$$

Hence

$$\underset{x,y,z}{\forall} \ x,y,z > 0 \implies (a+b+c)\,x + (b+c)\,y + cz \geq 0$$

Thus

$$a+b+c \geq 0 \qquad b+c \geq 0 \qquad c \geq 0$$

□

Lemma 5. *Let $R_{p,q}$ be a contracting quadratic map. Then we have*

$$\begin{array}{ll} p_3 + p_4 - 2 \geq 0 & p_4 - 1 \geq 0 \\ q_3 + q_4 - 2 \geq 0 & q_3 - 2 \geq 0 \end{array}$$

Proof. Let $R_{p,q}$ be a contracting quadratic map. Using Lemma 3, we can rewrite the condition (1) as

$$\underset{L,U,x}{\forall} \; 0 < L \leq \sqrt{x} \leq U \implies L \leq L + \frac{x - L^2}{p_3 L + p_4 U} \leq \sqrt{x} \leq U + \frac{x - U^2}{q_3 U + q_4 L} \leq U$$

Simplifying and splitting, we have

$$\begin{aligned} \underset{L,U,x}{\forall} \; 0 < L \leq \sqrt{x} \leq U &\implies 0 \leq \frac{(\sqrt{x} - L)(\sqrt{x} + L)}{p_3 L + p_4 U} \leq \sqrt{x} - L \\ \underset{L,U,x}{\forall} \; 0 < L \leq \sqrt{x} \leq U &\implies 0 \leq \frac{(U - \sqrt{x})(U + \sqrt{x})}{q_3 U + q_4 L} \leq U - \sqrt{x} \end{aligned}$$

By restricting the universal quantification to $\sqrt{x} \neq L$ and $\sqrt{x} \neq U$, we have

$$\begin{aligned} \underset{L,U,x}{\forall} \; 0 < L < \sqrt{x} < U &\implies 0 \leq \frac{\sqrt{x} + L}{p_3 L + p_4 U} \leq 1 \\ \underset{L,U,x}{\forall} \; 0 < L < \sqrt{x} < U &\implies 0 \leq \frac{\sqrt{x} + U}{q_3 U + q_4 L} \leq 1 \end{aligned}$$

By canceling the denominators, we have

$$\begin{aligned} \underset{L,U,x}{\forall} \; 0 < L < \sqrt{x} < U &\implies \sqrt{x} + L \leq p_3 L + p_4 U \\ \underset{L,U,x}{\forall} \; 0 < L < \sqrt{x} < U &\implies \sqrt{x} + U \leq q_3 U + q_4 L \end{aligned}$$

By rewriting it, we have

$$\begin{aligned} \underset{L,U,x}{\forall} \; 0 < L < \sqrt{x} < U &\implies (p_3 - 1) L - \sqrt{x} + p_4 U \geq 0 \\ \underset{L,U,x}{\forall} \; 0 < L < \sqrt{x} < U &\implies q_4 L - \sqrt{x} + (q_3 - 1) U \geq 0 \end{aligned}$$

From Lemma 4, we have

$$\begin{array}{lll} (p_3 - 1) + (-1) + (p_4) \geq 0 & (-1) + (p_4) \geq 0 & (p_4) \geq 0 \\ (q_4) + (-1) + (q_3 - 1) \geq 0 & (-1) + (q_3 - 1) \geq 0 & (q_3 - 1) \geq 0 \end{array}$$

Simplifying, we finally have

$$\begin{array}{ll} p_3 + p_4 - 2 \geq 0 & p_4 - 1 \geq 0 \\ q_3 + q_4 - 2 \geq 0 & q_3 - 2 \geq 0 \end{array}$$

□

Now we are ready to prove the two claims in Main Theorem. The following lemma (Lemma 6) will prove the claim (a) and the subsequent lemma (Lemma 7) will prove the claim (b).

Lemma 6 (Main Theorem (a))**.** *Let $R_{p,q}$ be a contracting quadratic map which is not R_{p^*,q^*} (Secant-Newton). Then we have*

$$\underset{L,U,x}{\forall} \quad 0 < L \le \sqrt{x} \le U \quad \Longrightarrow \quad R_{p^*,q^*}([L,U],x) \ \subseteq \ R_{p,q}([L,U],x)$$

Proof. Let $R_{p,q}$ be a contracting quadratic map which is not R_{p^*,q^*} (Secant-Newton), that is, $p \neq p^*$ or $q \neq q^*$. Let L, U, x be arbitrary such that $0 < L \le \sqrt{x} \le U$. We need to show

$$R_{p^*,q^*}([L,U],x) \ \subseteq \ R_{p,q}([L,U],x)$$

Note

$$\begin{aligned}
& R_{p^*,q^*}([L,U],x) \ \subseteq \ R_{p,q}([L,U],x) \\
\Longleftrightarrow\ & L' \le L^* \ \wedge \ U^* \le U' \\
\Longleftrightarrow\ & L + \tfrac{x-L^2}{p_3 L + p_4 U} \le L + \tfrac{x-L^2}{L+U} \\
& \wedge \qquad\qquad \text{(Due to Lemma 3)} \\
& U + \tfrac{x-U^2}{2U} \le U + \tfrac{x-U^2}{q_3 U + q_4 L} \\
\Longleftrightarrow\ & \left(x-L^2\right)\left(\tfrac{1}{L+U} - \tfrac{1}{p_3 L + p_4 U}\right) \ge 0 \\
& \wedge \\
& \left(U^2-x\right)\left(\tfrac{1}{2U} - \tfrac{1}{q_3 U + q_4 L}\right) \ge 0 \\
\Longleftrightarrow\ & \left(x-L^2\right)\left(\tfrac{1}{2L+(U-L)} - \tfrac{1}{(p_3+p_4)L + p_4(U-L)}\right) \ge 0 \\
& \wedge \\
& \left(U^2-x\right)\left(\tfrac{1}{2L+2(U-L)} - \tfrac{1}{(q_3+q_4)L + q_3(U-L)}\right) \ge 0 \\
\Longleftrightarrow\ & \left(x-L^2\right)\tfrac{(p_3+p_4-2)L+(p_4-1)(U-L)}{(2L+(U-L))((p_3+p_4)L+p_4(U-L))} \ge 0 \\
& \wedge \\
& \left(U^2-x\right)\tfrac{(q_3+q_4-2)L+(q_3-2)(U-L)}{(2L+2(U-L))((q_3+q_4)L+q_3(U-L))} \ge 0 \\
\Longleftrightarrow\ & \left(x-L^2\right)\left((p_3+p_4-2)L + (p_4-1)\left(U-L\right)\right) \ge 0 \\
& \wedge \qquad\qquad \text{(Due to Lemma 5)} \\
& \left(U^2-x\right)\left((q_3+q_4-2)L + (q_3-2)\left(U-L\right)\right) \ge 0 \\
\Longleftrightarrow\ & \mathit{true} \quad \text{(Due to Lemma 5)}
\end{aligned}$$

Main Theorem (a) has been proved. □

Lemma 7 (Main Theorem (b))**.** *Let $R_{p,q}$ be a contracting quadratic map which is not R_{p^*,q^*} (Secant-Newton). Then we have*

$$\underset{L,U,x}{\forall} \quad 0 < L < \sqrt{x} < U \quad \Longrightarrow \quad R_{p^*,q^*}([L,U],x) \ \subsetneq \ R_{p,q}([L,U],x)$$

Proof. Let $R_{p,q}$ be a contracting quadratic map which is not R_{p^*,q^*} (Secant-Newton), that is, $p \neq p^*$ or $q \neq q^*$. Let L, U, x be arbitrary such that $0 < L < \sqrt{x} < U$. We need to show

$$R_{p^*,q^*}([L,U],x) \ \subsetneq \ R_{p,q}([L,U],x)$$

Following a similar process as in the proof of Lemma 6, we have

$$
\begin{array}{ll}
 & R_{p^*,q^*}([L,U],x) \subsetneq R_{p,q}([L,U],x) \\
\Longleftrightarrow & L' < L^* \ \vee \ U^* < U' \quad \text{(Due to Lemma 6)} \\
\Longleftrightarrow & L + \frac{x-L^2}{p_3 L + p_4 U} < L + \frac{x-L^2}{L+U} \\
 & \vee \qquad\qquad\qquad\qquad \text{(Due to Lemma 3)} \\
 & U + \frac{x-U^2}{2U} < U + \frac{x-U^2}{q_3 U + q_4 L} \\
\Longleftrightarrow & \frac{1}{L+U} - \frac{1}{p_3 L + p_4 U} > 0 \\
 & \vee \qquad\qquad\qquad\qquad \text{(Since } L < \sqrt{x} < U\text{)} \\
 & \frac{1}{2U} - \frac{1}{q_3 U + q_4 L} > 0 \\
\Longleftrightarrow & \frac{1}{2L+(U-L)} - \frac{1}{(p_3+p_4)L + p_4(U-L)} > 0 \\
 & \vee \\
 & \frac{1}{2L+2(U-L)} - \frac{1}{(q_3+q_4)L + q_3(U-L)} > 0 \\
\Longleftrightarrow & \frac{(p_3+p_4-2)L + (p_4-1)(U-L)}{(2L+(U-L))((p_3+p_4)L + p_4(U-L))} > 0 \\
 & \vee \\
 & \frac{(q_3+q_4-2)L + (q_3-2)(U-L)}{(2L+2(U-L))((q_3+q_4)L + q_3(U-L))} > 0 \\
\Longleftrightarrow & (p_3+p_4-2)L + (p_4-1)(U-L) > 0 \\
 & \vee \qquad\qquad\qquad\qquad \text{(Due to Lemma 5)} \\
 & (q_3+q_4-2)L + (q_3-2)(U-L) > 0 \\
\Longleftrightarrow & p_3+p_4-2 \neq 0 \ \vee \ p_4 - 1 \neq 0 \\
 & \vee \qquad\qquad\qquad\qquad \text{(Due to Lemma 5)} \\
 & q_3+q_4-2 \neq 0 \ \vee \ q_3 - 2 \neq 0 \\
\Longleftrightarrow & \neg(p_3+p_4-2=0 \ \wedge \ p_4-1=0 \ \wedge q_3+q_4-2=0 \ \wedge \ q_3-2=0) \\
\Longleftrightarrow & \neg(p_3=1 \ \wedge \ p_4=1 \ \wedge q_4=0 \ \wedge \ q_3=2) \\
\Longleftrightarrow & \neg(p=p^* \wedge q=q^*) \quad \text{(Due to Lemma 2)} \\
\Longleftrightarrow & p \neq p^* \vee q \neq q^* \\
\Longleftrightarrow & \mathit{true}
\end{array}
$$

Main Theorem (b) has been proved. □

4 Conclusion

In this paper we investigated optimal methods for real square root computation by interval refining. More exactly, we proved that the well known Secant-Newton refinement map is the optimal among its natural generalizations, that is, among the maps that are contracting and are certain rational functions. This result motivates several interesting further questions.

- What about n-th root? It is natural to generalize the family of contracting quadratic maps to contracting degree n maps, that is, rational functions whose numerators are n-degree forms and whose denominators are $(n-1)$-degree forms. One asks what is the optimal map among the family of maps.
- What about dropping the condition "contracting"? Recall that Secant-Newton map is a particular instance of interval Newton map with slope where m is chosen to be U (footnote 1). If one chooses a different m value (from U), then

the interval Newton map with slope is not contracting. In practice, one remedies this by intersecting the result of the map with $[L, U]$ before the next iteration. This trivially ensures that the resulting map is contracting. This motivates a larger family of maps where a map is defined as a quadratic map composed with intersection with $[L, U]$. Again, one asks what is the optimal map among the larger family of maps.

We leave them as open problems/challenges for future research.

References

[1] G. Alefeld and J. Herzberger. *Introduction to Interval Computations*. Academic Press, Inc., New York, NY, 1983.

[2] N. Beebe. Accurate square root computation. Technical report, Center for Scientific Computing, Department of Mathematics,University of Utah, 1991.

[3] W. Cody and W. Waite. *Software Manual for the Elementary Functions*. Prentice-Hall, Englewood Cliffs, NJ, 1980.

[4] D. Fowler and E. Robson. Square root approximations in old Babylonian mathematics: YBC 7289 in Context. *Historia Mathematica*, 25(4):366–378, 1998.

[5] J.F. Hart, E.W. Cheney, C.L. Lawson, H.J. Maehly, C.K. Mesztenyi, J.R. Rice, H.C. Thacher Jr., and C. Witzgall. *Computer Approximations*. John Wiley, 1968. Reprinted, E. Krieger Publishing Company (1978).

[6] J. R. Meggitt. Pseudo division and pseudo multiplication processes. *IBM Journal of Research and Development*, 6(2):210–226, 1962.

[7] R. E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.

[8] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2009.

[9] D. R. Morrison. A method for computing certain inverse functions. *Mathematical Tables and Other Aids to Computation*, 10(56):202–208, 1956.

[10] N. Revol. Interval Newton iteration in multiple precision for the univariate case. *Numerical Algorithms*, 34(2–4):417–426, 2003.

[11] J. H. Wensley. A class of non-analytical iterative processes. *The Computer Journal*, 1(4):163–167, 1959.