6 Newton–Raphson method again. Now from a more theoretical point of view

6.1 Convergence order of the Newton–Raphson method

Now I am well prepared to give a precise mathematical explanation why the Newton–Raphson method approaches the sought root so much faster (compare with, say, bisection method). But first I would like to start a very general definition.

Definition 6.1. *Let f be continuous. Relaxation uses the sequence*

$$x_{n+1} = x_n - \lambda f(x_n), \quad n = 0, 1, 2, \dots,$$

where λ is a parameter to be determined.

In the relaxation sequence above I have

$$g(x) = x - \lambda f(x),$$

and according to the discussion in the previous section I would like to minimize my derivative $|g'(x)|$ and make it as close as possible to 0 to increase the convergence rate. Here I have

$$g'(x) = 1 - \lambda f'(x),$$

so the optimal choice is

$$\lambda = \frac{1}{f'(x)},$$

and I end up with the familiar recurrent formula for the Newton–Raphson method

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{6.1}$$

For the following I will need to recall the notion of Taylor's formula.

Let $f\colon (a,b) \longrightarrow \mathbf{R}$ be $(p+1)$ times continuously differentiable function and let $c \in (a,b)$. Then it can be proved that for any $x \in (a,b)$

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \ldots + \frac{f^{(p)}(c)}{p!}(x-c)^p + \frac{f^{(p+1)}(\xi)}{(p+1)!}(x-c)^{p+1},$$

where ξ is some point between x and c. Constant c is called the center of Taylor's formula. The main profit here is that we replace our function f with "almost" a polynomial, which becomes an exactly a polynomial if I drop the final term in the formula (it is also called the remainder).

Now let $\hat{x}$ be a root of f, i.e., $f(\hat{x}) = 0$. Assuming that f is twice continuously differentiable function, I can write Taylor's formula with the center at the n-th iterate x_n:

$$0 = f(\hat{x}) = f(x_n) + f'(x_n)(\hat{x} - x_n) + \frac{f''(\xi)}{2!}(\hat{x} - x_n)^2,$$

Math 488/688: Numerical Analysis

e-mail: artem.novozhilov@ndsu.edu. Fall 2021

where ξ is between $\hat{x}$ and x_n.

From (6.1) I have that

$$f(x_n) = f'(x_n)(x_n - x_{n+1}),$$

which, together with Taylor's formula, leads to

$$x_{n+1} - \hat{x} = -\frac{f''(\xi)}{2f'(x_n)}(x_n - \hat{x})^2. \tag{6.2}$$

Passing to the absolute values and using the usual notation for the absolute errors $\varepsilon_n = |x_n - \hat{x}|$, I can rewrite (6.2) as

$$\varepsilon_{n+1} = \left|\frac{f''(\xi)}{f'(x_n)}\right| \varepsilon_n^2.$$

Now I will assume that $f'(x_n) \neq 0$ for all n. Since my function is two times continuously differentiable, there must be a closed interval $I = [\hat{x} - \delta, \hat{x} + \delta]$ for some $\delta > 0$ such that

$$\left|\frac{f''(y)}{f'(x)}\right| \leq M$$

for all $x, y \in I$.

In other words I get an estimate

$$\varepsilon_{n+1} \leq M\varepsilon_n^2,$$

which proves that Newton–Raphson method squares the error at each step, and hence has the second order of convergence.

Moreover, from the last formula

$$\begin{aligned} \varepsilon_{n+1} &\leq M\varepsilon_n^2 \leq M(M\varepsilon_{n-1}^2)^2 \leq \\ &\leq M(M^2(M\varepsilon_{n-2}^2)^2)^2 \leq \ldots \\ &\leq \frac{1}{M}(M\varepsilon_0)^{2^{n+1}}. \end{aligned}$$

For my sequence of iterates to converge I must have that $\varepsilon_n \to 0$, which will be true if $M\varepsilon_0 < 1$. Therefore, I proved

Theorem 6.2. *Let $f\colon [a, b] \longrightarrow \mathbf{R}$ be twice continuously differentiable function. Let $\hat{x} \in [a, b]$ be its root. Assume that $f'(x) \neq 0$ for all $x \in [a, b]$. Then there exists an interval $I = [\hat{x} - \delta, \hat{x} + \delta]$ around the root $\hat{x}$ such that for any $x_0 \in I$ the Newton-Raphson method will converge to $\hat{x}$ and the order of convergence is two.*

Moreover, δ can be chosen such that $M\varepsilon_0 < 1$, where $M = \max_{x,y\in I} |f''(x)/f'(y)|$ and $\varepsilon_0 = |x_0 - \hat{x}|$.

Remark 6.3. The order of convergence two is usually interpreted as "the number of correct digits in the answer is approximately doubled at each iteration." While this certainly can be true for many examples, statements like this should be always taken with a grain of salt. It is possible to come up with particular equations, where this rule (doubling the number of correct digits at each step) does not hold. Take, e.g., equation

$$f(x) =$$

and analyze it numerically.

Example 6.4. Let
$$f(x) = x^3 + x - 2,$$
and we need to find the root $\hat{x} = 1$ using the Newton–Raphson method. I check directly that the interval $[0, 2]$ brackets my root. I also have
$$\min_{x\in[0,2]} |f'(x)| = \min_{x\in[0,2]} |3x^2 + 1| = 1,$$
and
$$\max_{x\in[0,2]} |f''(x)| = \max_{x\in[0,2]} |6x| = 12,$$
hence I can choose my constant $M = 12/2 = 6$. Since I must have that $M|x_0 - \hat{x}| < 1$ I see that taking x_0 anywhere in $(1 - 1/6, 1 + 1/6)$ will result in the required inequality, and hence I showed that for any
$$x_0 \in (5/6, 7/6)$$
the Newton–Raphson method will converge.

Of course the analysis in this example is somewhat artificial since for a real task we would never knew what the root is exactly.

Theorem 6.2 gives quite a nice and detailed description of the *local* properties of the Newton–Raphson method. Can we supplement it with any discussion on global properties? In some cases yes. For instance, I will leave it as an exercise to convince yourself that if $f'(x) > 0$ and $f''(x) > 0$ for all $x \in [\hat{x}, c]$ then for any initial point $x_0 \in (\hat{x}, c]$ the Newton-Raphson method will converge to $\hat{x}$. Similarly, if $f'(x) > 0$ and $f''(x) < 0$ for all $x \in [c, \hat{x}]$ then for any $x_0 \in [c, \hat{x})$ the iterates will approach $\hat{x}$. In general, however, I cannot make such bold conclusions and the situation may become very complicated as the following example shows.

Example 6.5. Let $f(x) = x^3 - x$, that is this equation has three roots $\hat{x}_0 = 0, \hat{x}_{\pm 1} = \pm 1$. Let me rewrite this equation in a form suitable for simple iterates:
$$x = \frac{3x - x^3}{2} = g(x),$$
which now clearly has the same three fixed points. By using the usual analysis I will find that $\hat{x}_{\pm 1}$ are stable (i.e., there is at least small interval around them such that g is a contraction there) and $\hat{x}_0$ is unstable.

Let me define *the basin of attraction* $B(\hat{x})$ of the fixed point $\hat{x}$ as the set of all initial conditions that converge to $\hat{x}$:
$$B(\hat{x}) = \{x_0 \colon x_n \to \hat{x},\, x_{n+1} = g(x_n)\}\,.$$

By the graphical analysis (or in some other way) I can see that if I take $x_0 \in I_1 = (0, \sqrt{3})$ (note that $g(\sqrt{3}) = 0$) my simple iterates will converge to $\hat{x}_{+1}$. Moreover, since the interval $I_2 = (-2, \sqrt{3})$ is mapped to $(0, 1)$, the initial conditions from I_2 will also belong to the basin of attraction of $\hat{x}_{+1}$. Looking further, by a graphical analysis I see that there must be an interval I_3, which is mapped by g into I_2 and hence also is in the basin of attraction. And so on. In words, the basin of attraction of $\hat{x}_{+1}$ consists of the union of infinitely many non overlapping intervals whose lengthes quickly approach zero.

Similar picture can happen for any iterative method, including the Newton–Raphson method, and in general there is little hope to understand the global behavior of iterates.

6.2 Concluding comments on solving scalar transcendental equations numerically