Solving nonlinear equations \((82.4, 2.1)\)

In the previous lecture, we showed an **iterative method** for computing the square root \(V a\), which only required elementary operations (e.g. addition, multiplication & division).

In summary, we had:

* Start by setting \(x_0 < \text{some initial guess for } V a\) >
* Generate the sequence \(x_0, x_1, \ldots, x_n, \ldots\) by:

\[
x_{n+1} = \frac{x_n^2 + a}{2x_n}
\]

We also showed the following 2 facts:

* If the sequence \(\{x_n\}\) converges, it will converge to a solution of \(x^2 - a = 0\) (i.e. \(x = \pm \sqrt{a}\))
* Assuming that for some \(k_0\), we have that

\[
\left| \frac{x_k - \sqrt{a}}{\sqrt{a}} \right| = "\text{small}" \quad \text{(for example, less than 1%)}
\]

then

\[
\left| \frac{x_k - \sqrt{a}}{\sqrt{a}} \right| \leq C \left| \frac{x_k - \sqrt{a}}{\sqrt{a}} \right|^2 \quad \text{for } k \geq k_0
\]

Thus, subsequent iterations **double** the correct significant digits.
This example is a special case of an algorithm for solving nonlinear equations, known as Newton's method (or, the Newton-Raphson method).

Here is the general idea:

If we "zoom" close enough to any smooth function, it looks more and more like a straight line (specifically, the "tangent" line).

The newton method suggests:

If, after \( n \) iterations we have approximated the solution of \( f(x) = 0 \) (a nonlinear equation) as \( x_n \), then:

- Form the tangent line at \( (x_n, f(x_n)) \)
- Select \( x_{n+1} \) as the intersection of the tangent line with the horizontal axis (\( y=0 \)).
If \((x_n, y_n) = (x_n, f(x_n))\) the tangent line to the plot of \(f(x)\) at \(x_n, y_n\) is

\[ y - y_n = \lambda (x - x_n) \]

where \(\lambda = f'(x_n)\) is the slope.

Thus

\[ y = y_n + \lambda (x - x_n) = y_n + f(x_n) + f'(x_n) (x - x_n) \]

If we set \(y = 0\), we obtain

\[ f(x_n) + f'(x_n) (x - x_n) = 0 \]

\[ x = x_n - \frac{f(x_n)}{f'(x_n)} \]

\[ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \]

Newton's method.
Consider our previous example. The square root $\sqrt{a}$ is the solution to the nonlinear equation $f(x) = x^2 - a = 0$.

Newton's method gives

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - a}{2x_n} = \frac{2x_n^2 - x_n^2 + a}{2x_n}$$

$$\Rightarrow x_{n+1} = \frac{x_n + a}{2x_n}$$

which is the method we considered previously.

A few comments about Newton's method $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$:

* It requires the function $f(x)$ to be not only continuous, but differentiable, as well. We will later see variants that do not explicitly require knowledge of $f'$ (in cases where the derivative may be hard to compute).

* If ever we have an approximation $x_n$ with $f'(x_n) \approx 0$, we should expect problems, especially if we are not close to a solution (we will be nearly dividing by 0). Graphically:

* If we don't start close to a solution, convergence may not be guaranteed (or it may take a large number of iterations).
Fixed point iteration (§2.1)

Newton's method is a special case for a broader category of methods for solving nonlinear equations, called fixed point iteration methods.

Generally, if $f(x) = 0$ is the nonlinear equation we seek to solve, a fixed point iteration method proceeds as follows:

$$x_0 = \text{initial guess}$$

$$x_{n+1} = g(x_n)$$

where $g(x)$ is a properly designed function.

Thus, we construct the sequence: $x_0, x_1, x_2, \ldots, x_k, \ldots$ which should ideally converge to a solution of $f(x) = 0$.

The following questions arise:

(i) If the iteration converges, does it converge to a solution of $f(x) = 0$?

(ii) Is the iteration guaranteed to converge?

(iii) How fast does the iteration converge (and (iv) When do we stop iterating?)?
Q1. If \( \{x_n\} \) converges, does it converge to a solution of \( f(x) = 0 \) ?

Taking limits on \( x_{n+1} = g(x_n) \), and assuming that \( \lim x_n = a \) and \( g \) is continuous, we get

\[
\lim x_{n+1} = \lim g(x_n)
\]

\[
a = g(a)
\]

The simplest way to guarantee that \( a \) is a solution to \( f(x) = 0 \) (i.e. \( f(a) = 0 \)) is if we can show that \( a = g(a) \) and \( f(a) = 0 \) are equivalent expressions.

There are more than one ways to make this happen, e.g.

\[
f(x) = 0 \iff x + f(x) = x \iff x = g(x), \text{ where } g(x) = x + f(x)
\]

or

\[
(x \to 0) \quad f(x) = 0 \iff e^x f(x) = 0 \iff e^x f(x) + x^2 = x^2 \iff \\
\quad e^{-x} f(x) + x^2
\]

\[
\frac{x}{x} = x \quad \text{or} \quad x = g(x), \quad g(x) = \frac{e^x f(x) + x^2}{x}
\]

or

\[
f(x) = 0 \iff -\frac{f(x)}{f'(x)} = 0 \iff x - \frac{f(x)}{f'(x)} = x \iff \\
\quad x = g(x) \text{ with } g(x) = x - \frac{f(x)}{f'(x)} \quad \text{Newton's method!}
\]
Unfortunately, simply "constructing" \( g(x) \) in a way that \( x=g(x) \iff f(x)=0 \) does not imply that the iteration will converge! Consider \( f(x) = x^2 - a = 0 \) (solution: \( \sqrt{a} \))

Simple formulations such as:

\[ x^2 = a \iff x = \frac{a}{x} = g(x) \ ? \]

\[ x_1 = \frac{a}{x_0}, \quad x_2 = \frac{a}{x_1} = \frac{a}{a/x_0} = x_0 \]

Thus the sequence alternates forever \( x_0, x_1, x_0, x_1, \ldots \)

or:

\[ x^2 - a + x = x \]

\[ g(x) \]

Let \( x_0 = 3, \ a = 10 \)

\[ x_1 = x_0^2 - 10 + x_0 = 9 - 10 + 3 = 2 \]
\[ x_2 = x_1^2 - 10 + x_1 = 4 - 10 + 2 = -4 \]
\[ x_3 = x_2^2 - 10 + x_2 = 16 - 10 - 4 = 2 \quad \text{repeats} \]

or Let \( x_0 = 5, \ a = 10 \) (bad initial guess)

\[ x_1 = 25 - 10 + 5 = 20 \]
\[ x_2 = 400 - 10 + 20 = 410 \]

"diverges to \( \infty \)!"

It is even likely that for some iterations \( x=g(x) \), the sequence will diverge regardless of how good the initial guess is (other than absolutely correct!)
Fortunately, there are ways to ensure \( \{x_n\} \) converges, for certain choices of \( g(x) \).

**Definition:** A function \( g(x) \) is a contraction in the interval \([a,b]\), if

\[
|g(x) - g(y)| \leq L |x - y|
\]

for any \( x, y \in [a,b] \) and \( L \leq [0,1) \).

**Examples:** \( g(x) = \frac{x}{2} \) : \( |g(x) - g(y)| = \frac{1}{2} |x - y| \quad \) for any \( x, y \).

\( g(x) = x^2 \) in \([0.1, 0.2]\)!

\[
|g(x) - g(y)| = |x^2 - y^2| = |x+y||x-y| \leq 0.3 |x-y|
\]

in this case we really needed \( x, y \in [a,b] \)!

With a contraction, we can show the following:

* Let \( a \) be the real solution of \( f(x) = 0 \), and assume \( |x_0 - a| < \delta \) (\( \delta \) = some positive number).

If \( g \) is a contraction on \((a-\delta, a+\delta)\), the fixed point iteration converges to \( a \)!
Proof: Since \( a \) is a solution, we have \( g(a) = a \).

\[
| x_1 - a | = | g(x_0) - g(a) | \leq L | x_0 - a | < L \delta \\
| x_2 - a | = | g(x_1) - g(a) | \leq L | x_1 - a | < L^2 \delta \\
\vdots \\
| x_k - a | < L^k \delta
\]

since \( L < 1 \) we have \( \lim_{k \to \infty} | x_k - a | = 0 \)

i.e. \( x_k \to a \).

Note: For simplicity we assumed that a value \( a \) such that \( g(a) = a \) does exist. However if \( g \) is a contraction we can more generally show that the fixed point iteration will converge, and the limit of course will satisfy \( g(a) = a \).

In some cases, it can be a lot of work to show that \( g \) is a contraction, using the definition. However, if we can compute the derivative \( g' \) we have a simpler criterion:

* If \( g \) is differentiable, and \( |g'(x)| \leq L \) (where \( 0 \leq L < 1 \)) in \([a, b] \), then \( g \) is a contraction on \([a, b] \).
Proof Let \( x, y \in [a, b] \) (with \( x \neq y \)).

The mean value theorem states that
\[
\frac{g(x) - g(y)}{x - y} = g'(\xi) \quad \text{for some } \xi \in (x, y)
\]

if \( |g'(\xi)| \leq L \), then
\[
\left| \frac{g(x) - g(y)}{x - y} \right| \leq L \Rightarrow |g(x) - g(y)| \leq L |x - y|
\]

Examples: \( g(x) = \sin \left( \frac{2x}{3} \right) \)
\[
|g'(x)| = \frac{2}{3} |\cos \left( \frac{2x}{3} \right)| \leq \frac{2}{3} < 1
\]

Let’s try to apply the derivative criterion, to see if
\[ g(x) = x - \frac{f(x)}{f'(x)} \quad \text{(from Newton’s method)} \]
\[ \text{is a contraction!} \]

\[
g'(x) = 1 - \frac{f'(x)f(x) - f(x)f'(x)}{[f'(x)]^2} = 1 - 1 + \frac{f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2}
\]
Now, let us assume that:

- \( f(a) = 0 \) (i.e., \( a \) is a solution of \( f(x) = 0 \))
- \( f'(a) \neq 0 \)
- \( f'' \) is bounded near \( a \) (for example, if it is continuous).

Then

\[
\lim_{x \to a} g'(x) = \frac{f(x) f''(x)}{(f'(x))^2} = 0
\]

This means that there is an interval \( (a-\delta, a+\delta) \) where \( |g'(x)| \) is small (since \( \lim g'(x) = 0 \)).

Or specifically

\[|g'(x)| \leq L \quad (L < 1)\]

This means that \( g \) is a contraction on \( (a-\delta, a+\delta) \) and, if \( x_0 \in (a-\delta, a+\delta) \), the iteration will converge to the solution \( a \).