Math 128A, Summer 2019
Lecture 6, Tuesday 7/2/2019

Today's Agenda:
- Parameters
- fl (Fixed point iteration)
- Quadratic Convergence

1 On Parameters and Function Handles

```

[r,h] = bisection(a,b, f)

  y = bisection(a,b, @f, p)

function y = f(x,p)
  y = (x - eps ** 3) ** 3
  or  if  p == 0
          y = 0
      if  p == 1
          y = 1
      if  p > 1
          y = f(x, p-1) + f(x, p - 2)
  end
```

We call `@f` a "function handle". We can use parameters in the above way as `p`. We can pass through optional parameters.
For Homework 1, Problem 6, we need to
- (Write and) Print a program
- Print Results (save into a text file; matlab: `diary`, screenshot, or picture)
- Include Plots

2 Floating Point: Fixed Point Iteration

[1] Solve $f(x) = 0 \iff x = g(x)$. One obvious way is to add x to both sides of the equation, but it turns out this is usually a very bad idea.
[2] Come up with better ways to create $g(x)$.
[3] Then, iterate $x_{n+1} = g(x_n)$ until satisfied (the exact number of digits; within tolerance) of the exact solution $x = g(x)$. That is, we stop when

$$\frac{|x_n - g(x_n)|}{|x_n|} \leq \varepsilon \qquad \text{if } x_n \neq 0$$

Aside: In some institutions, our tolerance (RHS) is user-specified. But for us, we say anything user-specified is bad, so we use the absolute ε.

Yesterday, we proved the following theorem:

Theorem 2.1. If $a \leq x \leq b$ implies:

1. $a \leq g(x) \leq b$ [Invariant]
2. $|g'(x)| \leq \frac{1}{2}$ [Contractive],

then for any $x_n \in [a, b]$, we have:

$$x_{n+1} = g(x_n); x = g(x) \implies |x_n - x| \leq 2^{-x}|x_0 - x|$$

Moreover, this gives a **unique** solution.

For some insight, consider:

$$\begin{aligned} x_{n+1} - x = |e_{n+1}| &= |g'(\xi_n)e_n| \\ &\leq \frac{1}{2}|e_n| \\ &\vdots \\ &2^{-(n+1)}|e_0| \end{aligned}$$

Example: $\sqrt{\textbf{something}}$: It turns out that in our example $\sqrt{}$, we can get quadratic convergence, because as we get closer to our solution, we do so at an increasing rate.

$$\begin{aligned} \sqrt{}: \qquad g(x) &= \frac{1}{2}\left(x + \frac{a}{x}\right), \qquad \text{if } 1 \leq a \leq 2 \\ g'(x) &= \frac{1}{2}\left(1 - \frac{a}{x^2}\right) = 0, \qquad \text{if } x = \sqrt{a}. \end{aligned}$$

2.1 Showing Uniqueness of Solution

Suppose $x = g(x)$ and $y = g(y)$. Note that we aren't supposing anything for contradiction; to show uniqueness, we simply show these have to be equal. Consider:

$$\begin{aligned} |x - y| &= |g(x) - g(y)| \\ &= |g'(\xi)(x - y)| \\ &\leq \frac{1}{2}|x - y| \end{aligned}$$

Additionally, let $z := |x - y|$.

$$0 \leq z \leq \frac{1}{2}z \implies \frac{1}{2}z \leq 0 \implies 0 \leq z \leq 0$$

Hence $z = 0$, and $|x - y| = 0 \implies x = y$.

2.2 What happens in Floating Point Arithmetic?

Let's say we're trying to solve $0 = f(x) \iff x = g(x)$. If we're lucky (if we program it well), in floating point, this will be $g(x)(1 + \delta)$. This is an idealistic forward error. In other words, we deliver a g that is fairly accurate. Otherwise, we can deliver some $g\left(x(1 + \delta)\right)$, which is an **idealistic backward error**.

That is, we have:

$$fl[g(x)] = \begin{cases} g(x)(1+\delta) & \text{idealistic forward error} \\ g[x(1+\delta)] & \text{idealistic backwards error} \end{cases}$$

For example, we have $g(x)(1+17\delta_1)+18\delta_2$ realistically. So we see that:

Fixed point iteration is (somewhat) self error-correcting.

Consider a set of operations stacked within g, and the algorithms are fixed. So we assume that it does the best possible. $y = fl(x_n)$.

$$\begin{aligned} y_{n+1} &= g[y_n](1+\delta_n) \\ x &= g(x) \\ e_{n+1} &= y_{n+1} - x = g[y_n](1+\delta_n) - g(x)(1+\delta_n) + g(x)\delta_n \\ &= (1+\delta_n)\left[g(y_n) - g(x)\right] + \underbrace{x}_{=g(x)}\delta_n \end{aligned}$$

Notice the bridges (tricks with g) we use in the above manipulations to get our result. Now we use the MVT, put absolute values on everything and proceed as usual.

$$|e_{n+1}| \leq \frac{1}{2}(1+\varepsilon)|e_n| + |x|\varepsilon$$

Notice the differences from our previous error expression. We should be looking at

$$r_n := \frac{|e_n|}{|x|}, \qquad \text{relative error in } x_n,$$

where

$$r_{n+1} \leq \frac{1+\varepsilon}{2} r_n + \varepsilon$$

This is very common where our error at one step is the error at the previous step times some constant, plus an added error (constant). We accordingly call this linear.
Let's call $\theta := \frac{1+\varepsilon}{2}$ (because it looks like a 0 and 1 interval, taking a nap, according to Strain).
Consider from the above, we have the following pattern:

$$\begin{aligned} r_{n+1} &\leq \theta r_n + \varepsilon \\ r_n &\leq \theta r_{n-1} + \varepsilon \\ r_{n+1} &\leq \theta\left(\theta r_{n-1} + \varepsilon\right) + \varepsilon \\ &\vdots \\ &\leq \theta^{n+1} r_0 + \underbrace{\left(\theta^n + \theta^{n-1} + \cdots + \theta + 1\right)}_{\leq 2}\varepsilon \end{aligned}$$

where

$$r_0 = \frac{|x_0 - x|}{|x|} \leq \frac{|b-a|}{\min\{|a|,|b|\}}.$$

We verify, $fl\left(\frac{1}{2}+\frac{\varepsilon}{2}\right)=\frac{1}{2}+\frac{\varepsilon}{2}$.
We know this bound ≤ 2 because:

$$1+\theta+\cdots+\theta^n=\frac{1-\theta^{n+1}}{1-\theta}\leq\frac{1}{1-\theta}$$

and

$$r_{n+1}\leq\underbrace{\theta^{n+1}r_0}_{=0}+\underbrace{\frac{1}{1-\theta}}\varepsilon\leq 2\varepsilon$$

We call the right term the "black hole" of error. In total, this is a very good result.
Recall that back when we were summing $\sum\frac{1}{k^2}$ from left → right, we had $O(n\varepsilon)$, and at best, $O(\sqrt{\varepsilon})$. So if we're using this summation to compute π, then we won't get good results. However, if we define π as the root of a fixed point iteration, then it's a whole different story (we can get good accuracy).
Recall

$$\varepsilon=2^{-52}\implies\sqrt{\varepsilon}=2^{-26}=0.000\cdots00100\cdots0,$$

as opposed to

$$2\epsilon=0.000\cdots0010.$$

3 Quadratic Convergence

We look at $\sqrt{x}$ as an example.

$$\begin{aligned}
g(x)&=\frac{1}{2}\left(x+\frac{a}{x}\right)\\
x_{n+1}&=g(x_n)\\
x&=g(x)\\
\implies x_{k+1}-x&=g(x_n)-g(x)\\
&=g'(\xi_n)(x_n-x)\quad\text{(MVT)}
\end{aligned}$$

With $\xi_n\in(x_n,x)\to x$, if $g'(x)=0$. To capture the fact that our error gets reduced, we use an additional term of the taylor expansion (linear MVT does not suffice).
So we look at the Taylor Remainder:

$$\begin{aligned}
g(x_n)-g(x)&=g(x)+g'(x)(x_n-x)+\frac{1}{2}g''(\xi_n)(x_n-x)^2-g(x)\\
&=g'(x)(x-x_n)+\frac{1}{2}g''(\xi_n)(x-x_n)^2\\
&=\frac{1}{2}g''(\xi_n)(x-x_n)^2\quad\text{(We note that } g'(x)=0\text{ when } x=g(x).)
\end{aligned}$$

So we have,

$$\begin{aligned}
e_{n+1}&=\frac{1}{2}g''(\xi_n)e_n^2,\qquad\text{or equivalently,}\\
&=\frac{1}{2}\left(g''(\xi_n)g_n\right)e_n
\end{aligned}$$

Example $\sqrt{4}$:

$$\begin{aligned}
x_0 &= 4\\
x_1 &= \frac{1}{2}\left(4+\frac{4}{4}\right) = 2.5 \quad \text{(error 0.5)}\\
x_2 &= \frac{1}{2}\left(2.6+\frac{4}{2.5}\right) = 2.05 \quad (\text{error } 0.05 = 5\times 10^{-2})\\
x_3 &= \frac{1}{2}\left(2.05+\frac{4}{2.05}\right) = 2.0061 \quad (\text{error } 6\times 10^{-4})\\
x_4 &= 2.0000000093 \quad (\text{error } 9\times 10^{-8})\\
x_5 &= 2.0 \quad (\text{error } 10\times 10^{-}16)
\end{aligned}$$

Consider

$$e_{n+1} := \frac{1}{2}g''(\xi_n)e_n^2.$$

If $a \le x \le b$, then $a \le g(x) \le b$ (Invariance),
and $|g''(x)| \le C, \qquad C|x_0 - x| \le 1$ (Contractive).

$$\begin{aligned}
|e_n| &\le \frac{1}{2}C|e_{n-1}|^2\\
\implies e_{n+1} &\le \frac{1}{2}C|e_n|^2 \le \frac{1}{2}C\left(\frac{1}{2}C|e_{n-1}|^2\right)^2\\
&= \left(\frac{1}{2}C\right)^3 \cdot |e_{n-1}|^4\\
&\le \left(\frac{1}{2}C\right)^3\left(\frac{1}{2}C|e_n-2|^2\right)^4\\
&= \left(\frac{1}{2}C\right)^7 |e_{n-1}|^8
\end{aligned}$$

Then,

$$\begin{aligned}
|e_n| &\le \left(\frac{1}{2}C\right)^{2-1}|e_0|^{2^n}\\
&= \frac{2}{C}\left(\frac{C|e_0|}{2}\right)^{2^n}
\end{aligned}$$

This is either great or terrible. It's wonderful if the expression in the parenthesis is close to $\frac{1}{2}$, and terrible for when that is close to 1 (will explode). So we say:

$$\begin{cases}\text{fast convergence} & \text{if } C|e_0| \le 1\\ \text{ divergence} & \text{if } C|e_0| \ge 2\end{cases}$$

$$\begin{aligned}
g'(x) &= \frac{1}{2}\left(1-\frac{a}{x^2}\right)\\
g''(x) &= \frac{1}{2}\frac{2a}{x^3} = \frac{a}{x^3}
\end{aligned}$$

So this gives

$$\left|\frac{a|x_0 - x|}{\min x^3}\right| \le 1 \to \checkmark \text{ iteration converges very fast}$$

From our theorem, we don't have anything for uniqueness, so we assume $|g'(x)| \le \frac{1}{2}$ for $a \le x \le b$, and $x = g(x) \implies g'(x) = 0$. We call this quadratic convergence because the error is small and being squared (quad for square).

Theorem 3.1. If $(a \le x \le b)$ implies:

1. $a \le g(x) \le b$,
2. $|g'(x)| \le \frac{1}{2}$,
3. $|g''(x)| \le C$, where $C|b - a| \le 1$
4. $x = g(x) \implies g'(x) = 0 \implies x_n \to x$ and

$$|x_{n+1} - x| \le \frac{1}{2} C |x_n - x|^2$$

3.1 How do we make $g'(x) = 0$ at $x = g(x)$?

Try $f(x) = 0$. Recall that g is what we design in trying to find roots to $f(x) = 0$. So there's some wiggle-room in how we define or design $g(x)$.

$$x = x + h(x)f(x) =: g(x)$$

And choose $h(x)$ such that $g'(x) = 0$ at the solution x. Differentiating, we get

$$g'(x) = 1 + h'(x) \underbrace{f(x)}_{0 \text{ at } x=0} + h(x)f'(x).$$

We want this $g'(x)$ to be 0 when $f(x) = 0$. So we just need $1 = hf'$, which says we should choose:

$$h(x) := \frac{-1}{f'(x)}.$$

So from

$$g(x) = x - \frac{f(x)}{f'(x)},$$

we have

$$\begin{aligned} g'(x) &= 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{[f'(x)]^2} \\ &= 1 - 1 + f(x)\frac{f''(x)}{f'(x)^2} \\ &= f(x)\frac{f''(x)}{[f'(x)]^2} = 0, \end{aligned}$$

if $f(x) = C$.

Remark: Be careful in trying to find roots of something like $f(x) = x^2$, where a slight perturbation of our data (slightly flawed or incorrect data) can make the zero cease to exist (if the parabola shifts up a very slight amount). Instead, we should be looking for a minimum of the function.

If $g(x) - x - \frac{f(x)}{f'(x)}$,

$$\begin{aligned} |g'(x)| &= \left|1 - \frac{f'(x)f'(x) - f(x)f''(x)}{[f'(x)]^2}\right| \\ &= \left|f(x)\frac{f''(x)}{[f'(x)]^2}\right| \leq \frac{1}{2}, \qquad \text{close enough to } x \ . \\ |g''(x)| &= f'\frac{f''(x)}{[f'(x)]^2} + ff'''\frac{1}{[f']^2} + ff''\frac{-2f''}{[f']^3} \end{aligned}$$

So if $\left|\frac{f''(x)}{f'(x)}\right| \leq k$, then we are probably ok.

3.2 Alternate Derivation of Newton's Method

We can expand and operate.

$$\begin{aligned} f(x) &= 0 \\ f(x_n) &\neq 0 \end{aligned}$$

But maybe we can expand $f(x)$ about x_n:

$$\begin{aligned} f(x) &= f(x_n) + f'(x_n)\underbrace{(x - x_n)}_{\text{solve for}} + \frac{1}{2}f''(\xi_n)(x - x_n)^2 \\ \implies x - x_n &= \frac{-f(x_n)}{f'(x_n)}\left(-\frac{1}{2}\frac{f''(\xi_n)}{f'(x_n)}(x - x_n)^2\right) \\ x &= x_n - \frac{f(x_n)}{f'(x_n)}\left(-\frac{1}{2}\frac{f''(\xi_n)}{f'(x_n)}(x - x_n)^2\right) \end{aligned}$$

This gives us Newton's method, but we didn't need to know anything from Fixed Point iteration.

Lecture ends here.