ME 360 Lecture 6

Now we know two methods of root finding, let's compare the two methods.

Note, the two things we care about are

1. accuracy 2. convergence speed.

Bisection Method

Since after each step of bisection method, we are dividing our interval in half. After n steps, the interval $[a_n, b_n]$ has length $(b-a)/2^n$.

After n steps, the estimated root $x_c = (a_n + b_n)/2$

Let r be the exact root, then we have

$$\text{solution error} = |x_c - r| < \frac{b-a}{2^{n+1}}$$

$$\text{Function evaluations} = n+2$$

In other words, for every function evaluation, we cut the uncertainty in the root by a factor of 2.

Definition of solution accuracy:

A solution is correct within p decimal places if the error is less than 0.5×10^{-p}.

Example: how many iterations do we need to have a solution that is correct within 6 decimal places for function $f(x) = \cos x - x$ in the interval $[0, 1]$?

Hint: $f(x)$ doesn't matter here!

using the relationship of solution error, we have

$$\frac{b-a}{2^{n+1}} < 0.5 \times 10^{-6}, \quad \text{here } b=1, a=0.$$

$$\Downarrow$$

$$\frac{1}{2^{n+1}} < 0.5 \times 10^{-6} \Rightarrow \quad n > 6/(\log_{10} 2) \approx 19.9$$

$\boxed{n = 20}$ $\therefore$ 20 iterations

Exercise in class:

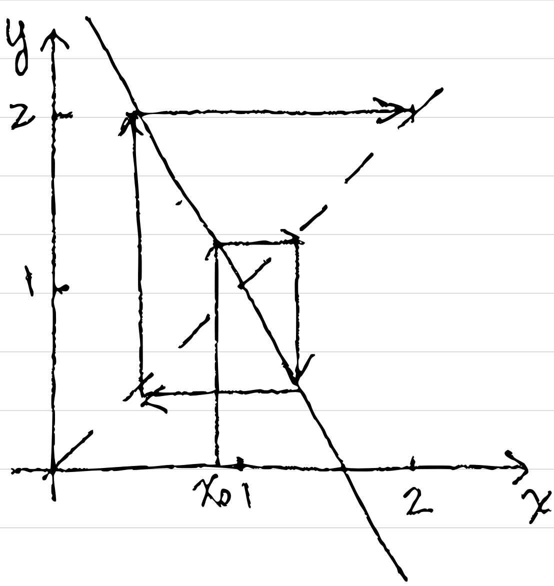
Linear Convergence of Fixed-Point Iteration

- How can we know when FPI will converge?
  Let's look at it using the simplest case possible: linear functions.

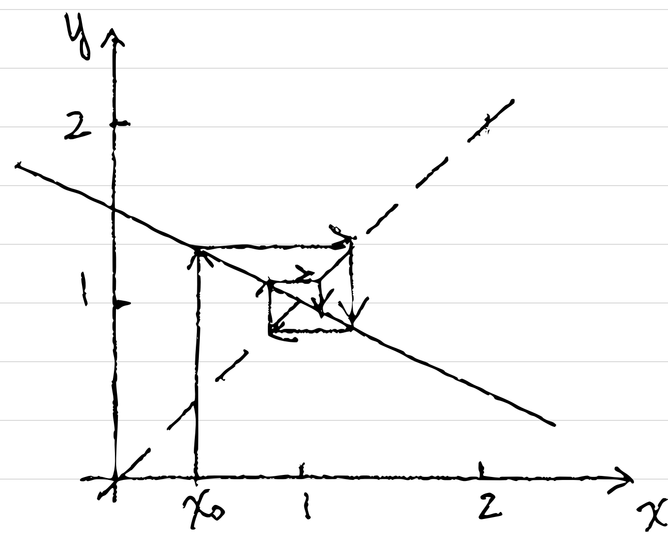
Two linear functions of interest:

$$g_1(x) = -\frac{3}{2}x + \frac{5}{2} \qquad g_2(x) = -\frac{1}{2}x + \frac{3}{2}$$

First, geometric argument.



Diverges



Converges

Let's now look at it from equations:
We want to rewrite the equations in the form of $(x-r)$ where r is the fixed point, in this case, $r=1$.

$$g_1(x) = -\frac{3}{2}(x-1) - \frac{3}{2} + \frac{5}{2} = -\frac{3}{2}(x-1)+1$$

$$\Downarrow$$

$$g_1(x)-1 = -\frac{3}{2}(x-1)$$

$$x_{i+1}-1 = -\frac{3}{2}(x_i-1)$$

Let $e_i = |r-x_i|$ being the error at step i, we see that, for function g_1, $\boxed{e_{i+1} = \frac{3}{2}e_i}$, meaning that every FPI step we take, the error increases by a factor of 3/2.

Similarly, we can study $g_2(x)$.

$$g_2(x) = -\frac{1}{2}(x-1) - \frac{1}{2} + \frac{3}{2} = -\frac{1}{2}(x-1)+1$$

$$g_2(x)-1 = -\frac{1}{2}(x-1)$$

$$x_{i+1}-1 = -\frac{1}{2}(x_i-1) \Rightarrow \boxed{e_{i+1} = \frac{1}{2}e_i}$$

For function g_2, every FPI step we take, the error decreases by a factor of 1/2.

Let's put everything together.

* Definition of linear convergence

Let e_i denote the error at step i of an iterative method.
If $\lim\limits_{i\to\infty} \frac{e_{i+1}}{e_i} = S < 1$,
the method is said to obey linear convergence with rate S.

Theorem on linear convergence of FPI

Assume that g is continuously differentiable, that $g(r)=r$, and that $S=|g'(r)|<1$. Then Fixed-Point Iteration converges linearly with rate S to the fixed point r for initial guesses sufficiently close to r.

Locally convergent

An iterative method is called locally convergent to r if the method converges to r for initial guesses sufficiently close to r.

Now let's use this theorem to look at the example from last class. The three functions were:

$g_1(x)=1-x^3 \qquad g_2(x)=(1-x)^{1/3} \qquad g_3(x)=\frac{1+2x^3}{1+3x^2}$

the fixed point is around 0.6823

1. $g_1'(x)=-3x^2 \qquad |g_1'(r)|=3\cdot(0.6823)^2\approx 1.3966>1$

2. $g_2'(x)=-\frac{1}{3}(1-x)^{-\frac{2}{3}} \qquad |g_2'(r)|=\frac{1}{3}(1-0.6823)^{-\frac{2}{3}}\approx 0.716<1$

3. $g_3'(x)=6x^2/(1+3x^2)-6x(1+2x^3)/(1+3x^2)^2$

$$=\frac{6x^2(1+3x^2)-6x(1+2x^3)}{(1+3x^2)^2}$$

$$=\frac{6x(x+3x^3-2x^3-1)}{(1+3x^2)^2}$$

$$=\frac{6x(x^3+x-1)}{(1+3x^2)^2}$$

$|g_3'(r)|=0$

By computing $S=|g'(r)|$, we see that $g_1(x)$ diverges, $g_2(x)$ converges slowly, $g_3(x)$ converges very fast, as $S=|g_3'(r)|=0$ is the smallest value S can get.

Stopping Criteria for FPI

Since we can't predict the number of steps we need for FPI, we need to have a stopping criteria. In general, there are three different ways:

For a predefined tolerance tol:

1. $|x_{i+1} - x_i| < tol$

2. $\dfrac{|x_{i+1} - x_i|}{|x_{i+1}|} < tol$ (solution not too close to 0)

3. $\dfrac{|x_{i+1} - x_i|}{\max(|x_{i+1}|, \theta)} < tol$, for some $\theta > 0$

Option 3 is a hybrid method of 1 & 2.

For an even more robust stopping criteria, we also want to add a maximum iteration steps, in case of non-convergence.

Convergence Speed of Bisection & FPI

Bisection	FPI
Guaranteed convergence	only locally convergent
$S = \frac{1}{2}$	$S = \lvert g'(r) \rvert$

FPI can be faster or slower than Bisection. To improve the performance of FPI, we will introduce Newton's method next, a particularly refined version of FPI, where S is designed to be zero.