CSC 338 Lecture 9

Problem: Find a root x^* of a smooth function $f(x)$, so $f(x^*)=0$.

Approach #1: Interval Bisection.

- Guaranteed convergence
- Rate of convergence independent of the function (can determine num iter required for a given absolute error tolerance)
- only evaluate $f(x)$ ~ no derivatives, ...
- slow to converge (linear)
- does not use much info about f, just its sign.
- require initial bracket.

Approach #2: Fixed Point Iter

$0 = x^3 - x - 1$

$x^3 = x+1$

$x = (x+1)^{1/3}$

eg// Find roots of $f(x) = x^3 - x - 1$

Fixed point iteration of $g(x) = x^3 - 1$ diverges

Fixed point iteration of $g(x) = (x+1)^{1/3}$ Converges

Convergence of Fixed-Point Iteration

Suppose $g(x)$ smooth has a fixed point $x^* = g(x^*)$.
If $|g'(x^*)| < 1$ then there is an interval containing x^* such that if x_0 is in that interval, then the fixed-point iteration $x_{k+1} = g(x_k)$ converges to x^*.

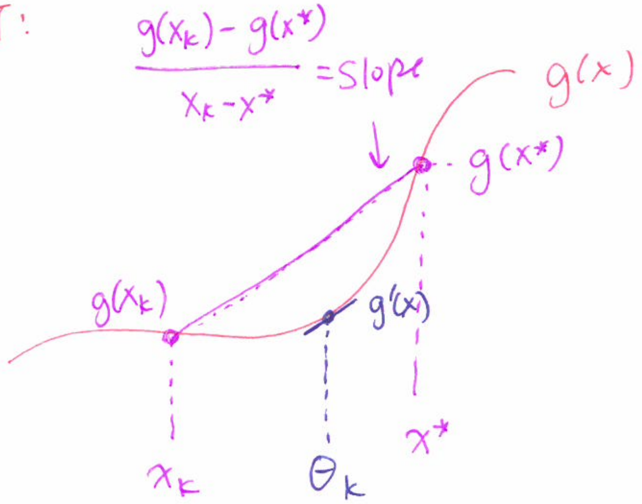
Pf (sketch) Suppose x^* is a fixed-point of $g(x)$

Then $e_{k+1} = x_{k+1} - x^*$

$= g(x_k) - g(x^*)$

Since $g(x)$ is smooth, we can apply the mean value thm.

MVT:



There exists θ_k between x_k and x^* such that

$$\frac{\overset{e_{k+1}}{g(x_k)-g(x^*)}}{\underset{e_k}{x_k-x^*}} = g'(\theta_k)$$

$$\Rightarrow \frac{e_{k+1}}{e_k} = g'(\theta_k)$$

If $|g'(x^*)| < 1$, then if we start x_0 close to x^*, then there exists a constant $c \in \mathbb{R}$ with $|g'(\theta_k)| \leq c < 1$

So $\displaystyle\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|} = \lim_{k\to\infty} |g'(\theta_k)| \leq c.$

So $|e_{k+1}| \leq c|e_k| \leq \dots \leq c^k |e_0|$

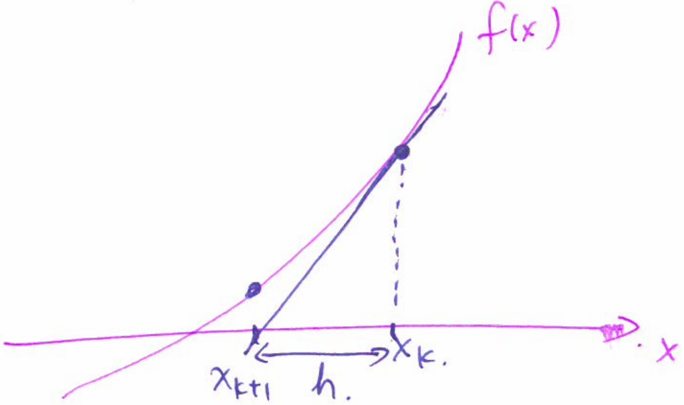
Since $c<1$, $|e_k| \to 0$ as $k \to \infty$

Fixed Point Iter:

- ↳ Converge faster than I.B. (if Fixed-Point Iter converges)
- ↳ Only need initial estimate.
- ↳ Might not converge.
- ↳ If $|g'(x)|$ is close to 1, then the convergence rate is slow
- ↳ We need a definition of $g(x)$ ⟷ hard to automate.

Newton's Method for Root Finding

Idea: Approximate f(x) with a linear function, and find ~~the~~ a root of the linear function



How do we approximate f(x) with a linear function at x_k?

⇒ Use the Taylor series approximation

$$f(x_k+h) = \underbrace{f(x_k) + f'(x_k)\cdot h}_{\text{linear approximation}} + O(h^2)$$

Set $0 = f(x_k) + f'(x_k)\cdot h$.

$\Rightarrow h = -\dfrac{f(x_k)}{f'(x_k)}$ $\Rightarrow x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$.

eg// $f(x) = x - 0.2\sin(x) - 0.5$.

k	x_k	$\lVert x_k - x_{k-1} \rVert$		e_k	
0	0.0.				
1	0.625	6.38×10^{-1}	$\times 2$	9.5×10^{-3}	$\times 2$
2	0.6154745	9.5×10^{-3}	$\times 2$	6.7×10^{-6}	$\times 2$.
3	0.6154682.	6.3×10^{-6}	$\times 2$	2.7×10^{-12}	
4	...	2.8×10^{-12}			

empirically, e_k decreases by a square factor each step

⇒ suggests quadratic convergence.

Newton's method is equivalent to finding a fixed-point of $g(x) = x - \frac{f(x)}{f'(x)}$, so to analyze the convergence of $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$ we can analyze

$$g'(x) = \cancel{1} - \frac{\cancel{f'(x) f'(x)} - f(x) f''(x)}{f'(x) f'(x)}$$

$$= \frac{f(x) f''(x)}{f'(x)^2}$$

So if $|g'(x)| < 1$, then Newton's method converges.

⇒ Newton's method converges when

$$\left| \frac{f(x) \cdot f''(x)}{f'(x)^2} \right| < 1.$$

can't be too far from root

curvature can't be too large

slope can't be too small

Break.

To measure the rate of convergence r of e_k, we need to find the largest value of r such that

$$\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|^r} = C \neq 0.$$

It is possible to show that $r=2$ for Newton's Method.

Newton's method:

- converges quickly, if convergent.
- don't need to specify $g(x)$, no initial bracket
  ~~Only require initial estimate~~
- need the derivative $f'(x)$.
- not guaranteed to converge.

Secant Method

One drawback of Newton's method is that we need to be able to evaluate both $\underline{f(x)}$ and $\underline{f'(x)}$.

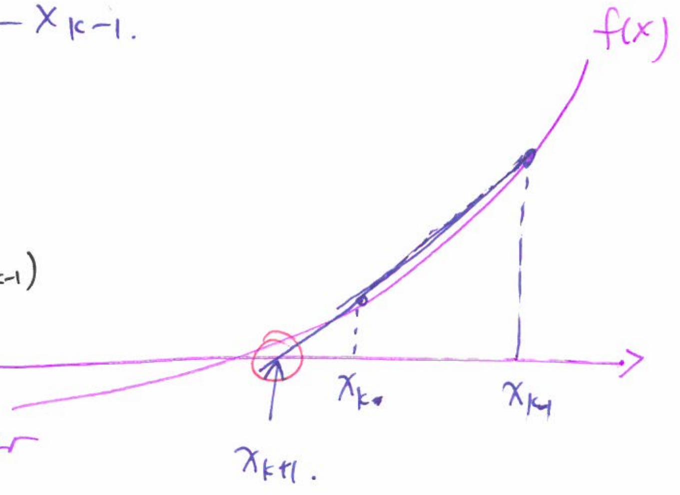
Idea: Use the successive iterate x_k, x_{k-1} to approximate $f'(x)$ using finite difference

$$f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}.$$

Algo: Start x_0, x_1

Take $x_{k+1} = x_k - f(x_k)\dfrac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})}$

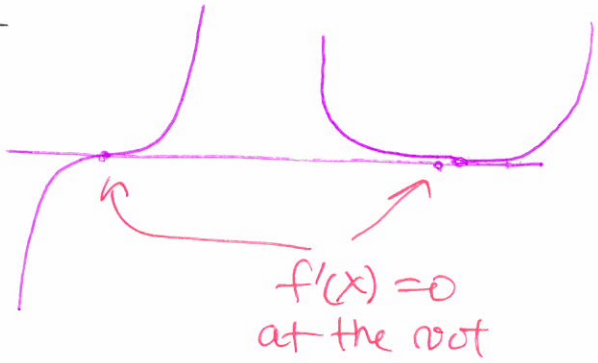
Can show that convergence rate of secant method is superlinear
($r = 1.618$)



Secant method is like Newton's Method, but requires no derivative

Multiple Roots

Def A solution x^* of $f(x) = 0$ is a root of multiplicity m if we can $f(x) = (x - x^*)^m q(x)$ where $q(x^*) \neq 0$.



Def If $m = 1$, x^* is called a simple root
$m > 1$, x^* is called a multiple root.

If $m > 1$, then the C.N. for finding the root x^* is $\left|\frac{1}{f'(x^*)}\right|$, which is infinite.

$\rightsquigarrow$ Newton's method will have a harder time finding multiple roots.

Robust Algo: which algo to implement?

In practice, combine a reliable but slow method like bisection with a fast but not always reliable method like secant method

New problem: Nonlinear optimization (§6).

Given a function $f:\mathbb{R}^n \to \mathbb{R}$ — objective function

and a set $S \subseteq \mathbb{R}^n$ — feasible set.

find $\underline{x}^* \in S$ such that f attains a minimum

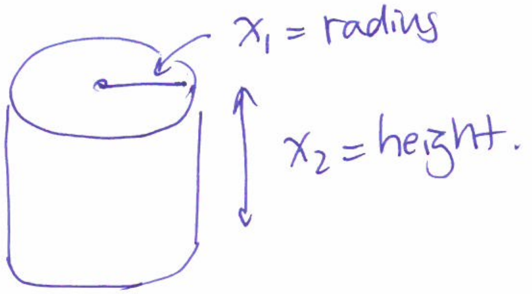
on S at x^*. i.e. $f(\underline{x}^*) \le f(\underline{x})$ for all $\underline{x} \in S$.

Normaly the set S is defined using constraints

$$S = \{\underline{x} \in \mathbb{R}^n \mid \underbrace{g(\underline{x}) = \underline{0}}_{\text{equality constraints}},\ \underbrace{h(\underline{x}) \le \underline{0}}_{\text{inequality constraints}}\}.$$

If $S = \mathbb{R}^n$, the optimization problem is unconstrained.

eg// minimize the surface area of a cylinder given a volume constraint



minimize :

$$f(x_1, x_2) = 2\pi x_1^2 + (2\pi x_1)x_2$$

Subject to:

$$g(x_1, x_2) = (\pi x_1^2)x_2 = 355$$

Local vs Global Minima

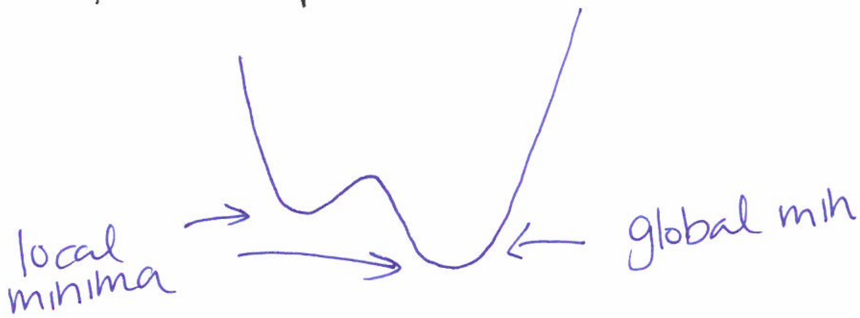
Def. The point $\underline{x}^*$ is a global minimum of f if $f(\underline{x}^*) \le f(\underline{x})$ for any $\underline{x} \in S$.

Def The point $\underline{x}^*$ is a local minimum of f if

$f(\underline{x}^*) \le f(\underline{x})$ within some neighborhood of $\underline{x}^*$

(there exists $\epsilon > 0$ such that if $\|\underline{x} - \underline{x}^*\| < \epsilon$
then $f(\underline{x}^*) \le f(\underline{x})$)

The techniques we discuss will be for finding local minima. It is generally difficult to find global minima unless f has special properties.



Unconstrained optimization — optimality condition.

In 1D: $f: \mathbb{R} \to \mathbb{R}$. a local minima always has:

① $f'(x) = 0$. ~ x is a critical point.

② $f''(x) > 0$.

Def A critical point of $f: \mathbb{R}^n \to \mathbb{R}$ is a point $\underline{x} \in \mathbb{R}^n$ such that the gradient of f is $\underline{0}$ at $\underline{x}$:

$$\nabla f(\underline{x}) := \begin{bmatrix} \frac{\partial f(\underline{x})}{\partial x_1} \\ \frac{\partial f(\underline{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\underline{x})}{\partial x_n} \end{bmatrix}$$

If $\underline{x}$ is a critical point, then $\nabla f(\underline{x}) = \underline{0}$.

eg// $f(\underline{x}) = x_1^2 + x_2^2$ $\quad f: \mathbb{R}^2 \rightarrow \mathbb{R}$.

$$\nabla f(\underline{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1}(x_1^2 + x_2^2) \\ \frac{\partial}{\partial x_2}(x_1^2 + x_2^2) \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$\Rightarrow$ There is a critical point at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

eg// $f(\underline{x}) = x_1^2 - x_2^2$.

$$\nabla f(\underline{x}) = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix}$$ Critical point at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Not all critical points are minima. Critical points can be maxima <u>or</u> saddle points.

We can distinguish between these by considering the Hessian matrix:

$$H_f(\underline{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$