Two problems:

(1) Given $F: \mathbb{R}^n \to \mathbb{R}^n$, find $x_* \in \mathbb{R}^n$ such that $F(x_*) = 0$

here $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\quad F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_n(x) \end{pmatrix}$

(2) Given $f: \mathbb{R}^n \to \mathbb{R}^1$ find

$x_* = \text{argmin } f(x)$ $\quad \to$ value of x that minimizes $f(x)$

will develop solution methods based on solving

$\nabla f(x) = 0$ $\qquad \nabla f(x) = \begin{pmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix}$

- Newton's method will (almost always) be our model method
- Will also consider (briefly) general fixed-point iteration.
- Will also look (briefly) at over-determined systems (nonlinear least-squares) and underdetermined systems

- Some realities:

(1) some systems have <u>no</u> solutions

$f(x) = \sin x + 2 = 0$

(2) some systems have many solutions

$f(x) = \sin x$

(3) solution method (almost always) <u>must</u> be iterative

<u>Topic 1 - Methods for single equations in one variable</u>

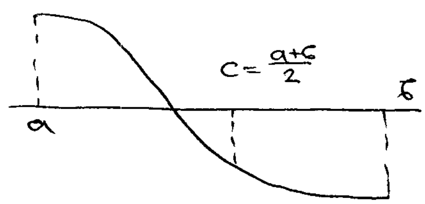
- first basic "pure" methods: bisection, Newton's method secant method

Bisection

Suppose f is continuous on $[a,b]$ and $f(a)\cdot f(b)<0$

Know: $\exists$ (at least one) $x_* \in [a,b]$ such that

$f(x_*)=0$ (IVT - intermediate value theorem)



Bisection method:

Given a, b such that $f(a)\cdot f(b)<0$

Iterate:

$$c = \frac{a+b}{2}$$

if $c=0$ stop

if $f(a)\cdot f(b)<0$, update $b \leftarrow c$

else update $a \leftarrow c$

need also to stop "when satisfied"

"when satisfied"?

Can show: if $\{[a_n, b_n], c_n\}$ are generated by bisection, then $\exists$ $x_* \in [a_0, b_0]$ such that $f(x_*)=0$ and $c_n \to x_*$ with

$$|c_n - x_*| \le \frac{b_0 - a_0}{2^{n+1}} \qquad (*)$$

provided f is continuous on $[a_0, b_0]$

Why? See $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ for each n

and $b_n - a_n \le \dfrac{b_0 - a_0}{2^n}$ so

(1) $\exists$ x_* such that $x_* \in [a_n, b_n]$ for every n

(2) $a_n \to x_*$ and $b_n \to x_*$

(3) $f(x_*) = \lim f(a_n) = \lim f(b_n)$ since f is continuous

(4) $0 \le f(x_*)^2 = \lim f(a_n) \cdot f(b_n) \le 0$ so $f(x_*) = 0$

So, if assumptions are met, have guaranteed convergence with bound $(*)$ on error

So given error tolerance $\varepsilon > 0$, can find n such that

$$\frac{b_0 - a_0}{2^{n+1}} \le \varepsilon$$

and know that $|c_n - x_*| \le \varepsilon$ after n steps

Newton's method

Suppose f is continuously differentiable near x_* such that $f(x_*) = 0$, $f'(x_*) \ne 0$

For x near x_*

$$0 = f(x_*) = f(x) + f'(x)(x_* - x) + o(x_* - x)$$

where $o(x_* - x)$ means $\frac{o(x_* - x)}{x_* - x} \longrightarrow 0$ as $x \to x_*$

"little oh"

Rearrange:

$$x_* = x - \frac{f(x)}{f'(x)} + \underbrace{\frac{o(x_* - x)}{f'(x)}}_{= o(x_* - x)}$$

So

$$x_+ = x - \frac{f(x)}{f'(x)} \quad \text{satisfies}$$

$$x_+ - x_* = o(x_* - x)$$

since $\frac{o(x_* - x)}{x_* - x} \longrightarrow 0$ as $x \to x_*$

expect that Newton's iterates $\{x_n\}$ converges to x_* with increasing speed

- suppose f is twice continuously differentiable near x_*

Then:

$$0 = f(x_*) = f(x) + f'(x)(x_* - x) + \frac{f''(\xi)}{2}(x_* - x)^2$$

$$\Rightarrow \quad x_* = x - \frac{f(x)}{f'(x)} - \frac{f''(\xi)}{2f'(x)}(x_* - x)^2$$

and see that Newton iterates $\{x_n\}$ converges to x_* with

$$|x_{n+1} - x_*| \leq \beta (x_n - x_*)^2$$

Can prove:

Theorem: Suppose f is continuously differentiable near x_* such that $f(x_*) = 0$ and $f'(x_*) \neq 0$. If x_0 is sufficiently near x_*, then Newton iterates $\{x_n\}$ converge to x_* with

$$\left.\begin{array}{l} |x_{n+1} - x_*| \leq c_n |x_n - x_*| \\ \text{where } c_n \to 0 \end{array}\right\} \quad (a)$$

- If f' also satisfies

$$|f'(x) - f'(x_*)| \leq L(x - x_*) \quad \ldots (b)$$

for x near x_* then

$$|x_{n+1} - x_*| \leq \beta |x_n - x_*|^2 \quad \ldots (c)$$

for β independent of n

- If f is <u>twice</u> continuously differentiable near x_* and $x_n \neq x_*$ for all n, then

$$\frac{x_{n+1} - x_*}{(x_n - x_*)^2} = -\frac{f''(x_*)}{2f'(x_*)} \quad \ldots (d)$$

Remarks:

(d) is just a nice thing for one equation in one unknown

(c) is quadratic convergence

This is (ultimately) <u>very</u> fast

Equivalent to

$$\left(\beta |x_{n+1} - x_*|\right) \leq \left(\beta |x_n - x_*|\right)^2$$

so $\beta \approx 1$ $\Rightarrow$ approximately double number of significant digits at each iteration

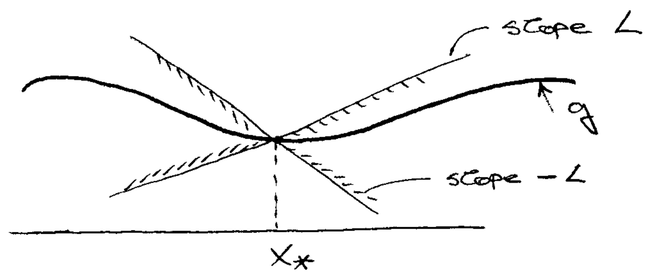
(Major strength of Newton's method)

(b) is called Lipschitz continuity of f'

Def: g is Lipschitz continuous at x_* if there exists an L such that

$$|g(x) - g(x_*)| \leq L|x - x_*|$$

for all x near x_*



Claim:

g differentiable at x_* $\overset{\not\Leftarrow}{\Rightarrow}$ g Lipschitz continuous at x_* $\overset{\not\Leftarrow}{\Rightarrow}$
(i) (ii)

$\overset{\not\Leftarrow}{\Rightarrow}$ g continuous at x_*
(iii)

(i) ⇒ (iii) is immediate since

$$|g(x) - g(x_*)| \le L|x - x_*|$$

$$\Rightarrow \lim_{x \to x_*} g(x) = g(x_*)$$

(i) ⇒ (ii) If g is differentiable at x_*

$$\Rightarrow g(x) = g(x_*) + g'(x_*)(x - x_*) + \theta(x - x_*)$$

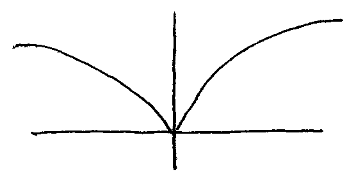
$$\Rightarrow |g(x) - g(x_*)| \le \left| g'(x_*) + \frac{\theta(x-x_*)}{x-x_*} \right| |x - x_*|$$

$$\le L|x - x_*|$$

for some L and x near x_*

(iii) ⇏ (ii)

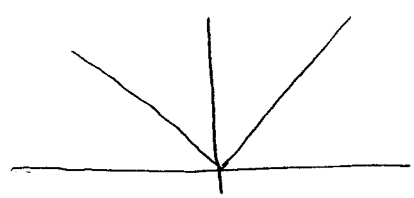
$$g(x) = |x|^{1/2}$$



continuous at $x_* = 0$ but not Lipschitz continuous there

(ii) ⇏ (i)

$$g(x) = |x|$$



Lipschitz continuous at $x_* = 0$ but not differentiable

(a) is called <u>superlinear</u> convergence

Def: suppose $x_n \to x_*$. Say convergence is

(i) linear if $\exists\, \lambda \in [0,1)$ such that

$$|x_{n+1} - x_*| \leq \lambda\, |x_n - x_*|$$

(ii) superlinear if $\exists\, \{\gamma_n\}$ such that $\gamma_n \to 0$ and

$$|x_{n+1} - x_*| \leq \gamma_n |x_n - x_*|$$

iii) order p, $1 < p < \infty$ if $\exists\, \beta$ such that

$$|x_{n+1} - x_*| \leq \beta\, |x_n - x_*|^p$$

$p = 2$ ~ quadratic convergence
$p = 3$ ~ cubic convergence

strictly speaking, these are called:

q - linear
q - superlinear
q - order p convergence

(q - quotient)

There is also r - convergence (r - root)

Say $x_n \to x_*$, r - linearly; r - superlinearly
r - order p for $1 < p < \infty$

if $|x_n - x_*| \leq \beta_n$ for each n and $\beta_n \to 0$

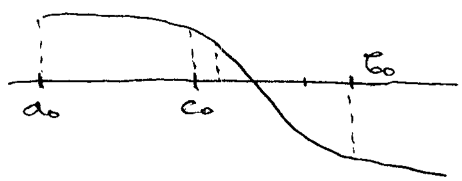
q - linearly
q - superlinearly
q - order p for $1 < p < \infty$

From now on, convergence is q - convergence unless explicitly noted.

Note: for bisection

$$|c_n - x_*| \leq \frac{b_0 - a_0}{2^{n+1}}$$

Follows that $c_n \to x_*$ r-linearly
(but convergence isn't q-linear in general)



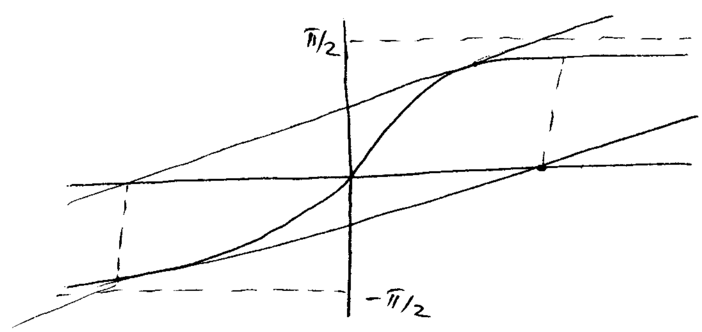
c_1 is further away from x_* than c_0

r-convergence, you are bounded by

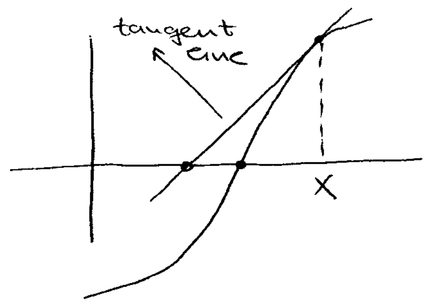
q-convergence - error is smaller and smaller

Convergence of Newton iteration is only <u>local</u>
i.e. guaranteed only for x_0 sufficiently near x_*

<u>Example</u>:
$f(x) = \tan^{-1} x$



General picture how Newton's method works



$$x_+ = x_0 - \frac{f(x)}{f'(x)}$$

Compare: (all we need to get method going)

Bisection: - need $[a_0, b_0]$ such that $f(a) \cdot f(b) < 0$

- guaranteed convergence if f continuous on $[a_0, b_0]$ with $|c_n - x_*| \leq \frac{b_0 - a_0}{2^{n+1}}$

- r-linear (slow)

- each iteration requires one f-evaluation

Newton: - can start with any x

- iterates may diverge if x_0 not near solution

- if $x_n \to x_*$ convergence is usually quadratic (very fast)

- each iteration requires evaluation of f' and f

- convergence is only local

More considerations for Newton

When to stop is a serious issue

Questions?

(1) Have we solve the problem

(2) Have we bogged down

(3) Have we run out of time/patience/money?

for (3) prescribe some ITMAX

test: $n = \text{ITMAX} \Rightarrow$ STOP

for (1) prescribe some ε_f

test: $|f(x_n)| \leq \varepsilon_f \Rightarrow$ STOP

Caution: must choose ε_f with an eye towards the scale of f

for (2) prescribe some ε_x

test: $|x_{n+1} - x_n| \leq \varepsilon_x \Rightarrow$ STOP

Similar caution:

For Newton

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

$$= -\frac{\overset{0}{\cancel{f(x_n)}} + f'(x_n)(x_n - x_*) + \overset{\text{small}}{O(x_n - x_*)}}{\underset{\approx f'(x_n)}{f'(x_n)}}$$

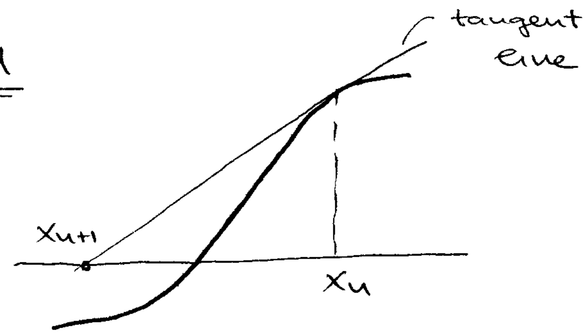
$$\approx -(x_n - x_*) \quad , \text{ for } x_n \text{ near } x_*$$

So this test is a test on $|x_n - x_*|$ where x_n is near x_*

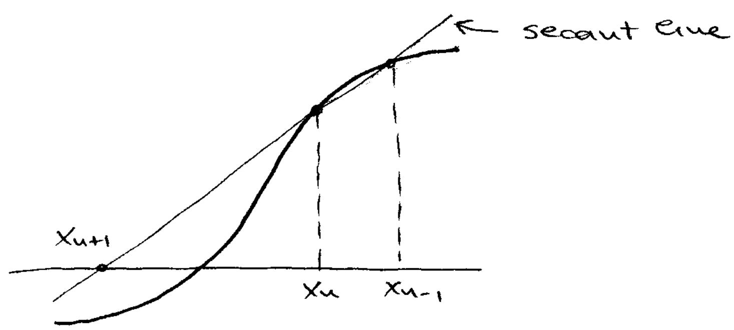
- In practice (1) is usually the primary test, ε_x is taken very small

Secant method

Newton



secant



get:

$$x_{n+1} = x_n - \left[\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}\right]^{-1} f(x_n)$$

Theorem: Suppose f is continuously differentiable near x_* such that $f(x_*) = 0$ and $f'(x_*) \neq 0$ then for x_0, x_{-1} sufficiently near x_* secant iterates $\{x_n\}$ converge to x_n with

$$\left\{ \begin{array}{l} |x_{n+1} - x_*| \leq c_n |x_n - x_*| \\ \text{where} \quad c_n \to 0 \end{array} \right\} \quad \text{superlinear}$$

If f is twice continuously differentiable near x_* and $x_n \neq x_*$ for any n, then

$$\lim_{n\to\infty} \frac{|x_{n+1} - x_*|}{|x_n - x_*|^{\frac{1+\sqrt{5}}{2}}} = \left| \frac{f''(x_*)}{2f'(x_*)} \right|^{\frac{1-\sqrt{5}}{2}}$$

i.e. convergence is order $\frac{1+\sqrt{5}}{2}$ (golden ratio)

q - convergence of order $\frac{1+\sqrt{5}}{2}$

Secant method:

- only one f-evaluation per iteration (don't need f')
- superlinear convergence (usually)
  - usually fast enough
- convergence only local
- for stopping, same test as for Newton's method

In practice, don't need to stick to "pure" methods

Q: Can we construct "hybrid" methods with advantages of each of the "pure" methods

Next time: Brent's algorithm implemented in MATLAB (fzero)