chapter 2 : rootfinding

section 2.1 : bisection method

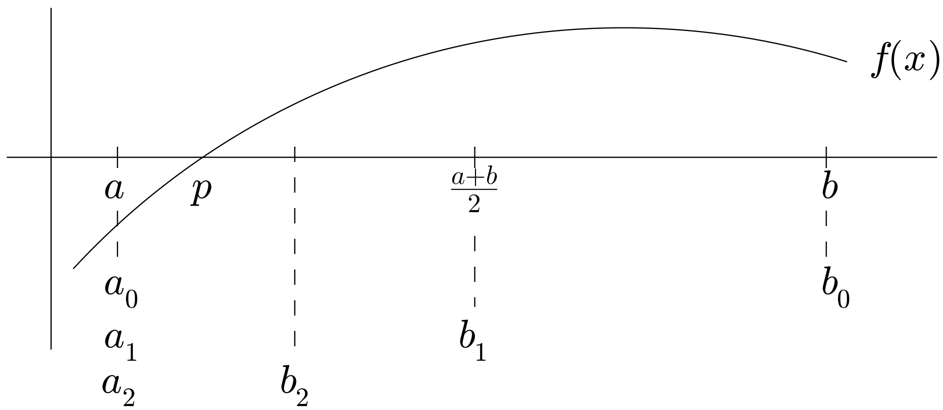
def : Given $f(x)$, a number p satisfying $f(p) = 0$ is called a root of $f(x)$.

ex : $f(x) = x^2 - 3x + 2 \Rightarrow p = 1, 2$

$f(x) = x^2 - 3 \Rightarrow p = \pm\sqrt{3}$

question : How can we find the roots of a general function $f(x)$?

idea : Find an interval $[a, b]$ such that $f(a)$ and $f(b)$ have opposite sign. Then $f(x)$ has a root in $[a, b]$ by the Intermediate Value Theorem (Math 451 - advanced calculus).



Consider the midpoint $\frac{a+b}{2}$. The root is contained in either the left subinterval $[a, \frac{a+b}{2}]$ or the right subinterval $[\frac{a+b}{2}, b]$; to determine which one, compute $f(\frac{a+b}{2})$. Then repeat.

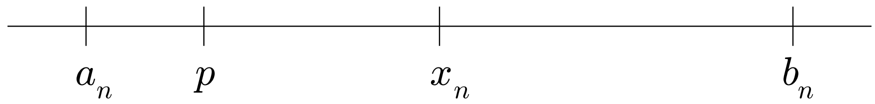
bisection method (assume $f(a) \cdot f(b) < 0$)

1. $n = 0$, $a_0 = a$, $b_0 = b$
2. $x_n = \frac{a_n+b_n}{2}$: current estimate of the root
3. if $f(x_n) \cdot f(a_n) < 0$, then $a_{n+1} = a_n$, $b_{n+1} = x_n$
4. else $a_{n+1} = x_n$, $b_{n+1} = b_n$
5. set $n = n + 1$ and go to line 2

ex : $f(x) = x^2 - 3$, $f(1) = -2$, $f(2) = 1 \Rightarrow$ there is a root p in $[1, 2]$, $p = 1.73205$

n	a_n	b_n	x_n	$f(x_n)$	$\lvert p - x_n \rvert$
0	1	2	1.5	-0.75	0.2321
1	1.5	2	1.75	0.0625	0.0179
2	1.5	1.75	1.625	-0.3594	0.1071
3	1.625	1.75	1.6875	-0.1523	0.0446
4	1.6875	1.75	1.71875	-0.0459	0.0133

<u>error bound</u> for the bisection method



$|p - x_n| \leq | b_n - a_n| = \frac{1}{2}|b_{n-1} - a_{n-1}| = (\frac{1}{2})^2|b_{n-2} - a_{n-2}| = \cdots = (\frac{1}{2})^n|b_0 - a_0|$

<u>ex</u> : how many steps are needed to ensure that the error is less than 10^{-3} ?

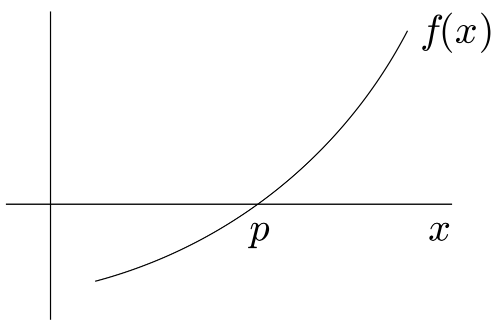
$(\frac{1}{2})^n|b - a| \leq 10^{-3} \;\Rightarrow\; n \geq 10$

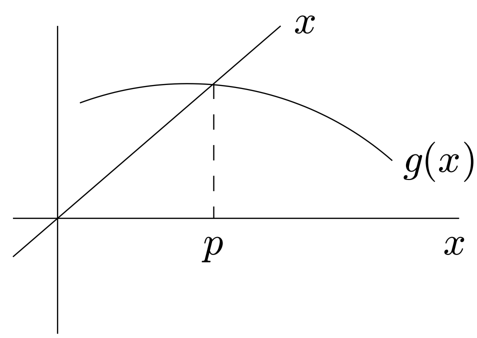
<u>stopping criterion</u> : here are three options

$|b_n - a_n| < \epsilon \;\;,\;\; |f(x_n)| < \epsilon \;\;,\;\; n = n_{\max}$

<u>section 2.3</u> : fixed-point iteration

Suppose that $f(x) = 0$ is equivalent to $x = g(x)$. Then p is a root of $f(x)$ if and only if p is a <u>fixed point</u> of $g(x)$.





<u>ex</u> : $f(x) = x^2 - 3 = 0$

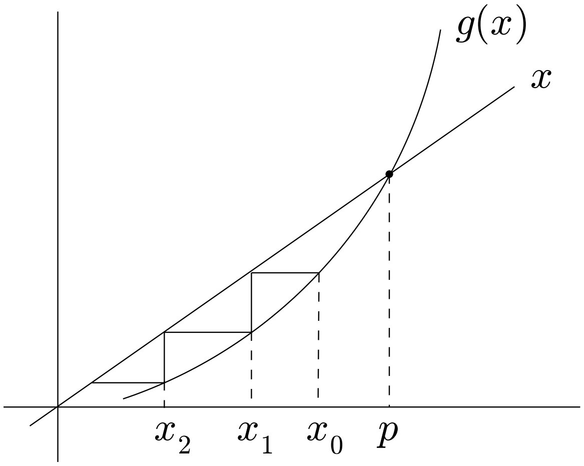
$$x = \frac{3}{x} = g_1(x) \;\;,\;\; x = x - (x^2 - 3) = g_2(x) \;\;,\;\; x = x - \left(\frac{x^2 - 3}{2}\right) = g_3(x)$$

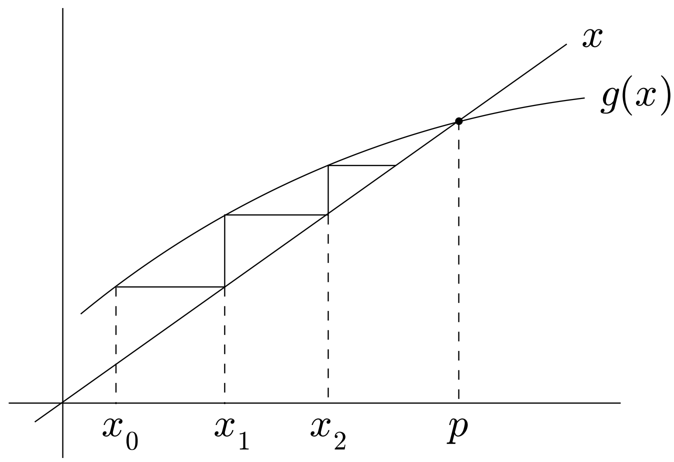
We try to solve $x = g(x)$ by computing $x_{n+1} = g(x_n)$ with some initial guess x_0. This process is called <u>fixed-point iteration</u>.

	case 1	case 2	case 3
n	x_n	x_n	x_n
0	1.5	1.5	1.5
1	2	2.25	1.875
2	1.5	0.1875	1.6172
3	2	3.1523	1.8095
4	1.5	-3.7849	1.6723
5	2	-15.1106	1.7740

Case 1 and case 2 diverge, but case 3 converges (recall: $p = 1.73205$).

question : what determines whether fixed point iteration converges or diverges? Let's consider two examples.





The 1st example diverges and the 2nd example converges.

thm

Let $k = \max|g'(x)|$. Then fixed-point iteration converges if and only if $k < 1$.

note : this is consistent with the two examples above.

pf

$$|p - x_{n+1}| = |g(p) - g(x_n)| \underset{\substack{\uparrow \\ \text{Mean Value Theorem}}}{=} |g'(\zeta)(p - x_n)| \leq k|p - x_n|$$

$$|p - x_{n+1}| \leq k|p - x_n| \leq k^2|p - x_{n-1}| \leq \cdots \leq k^{n+1}|p - x_0| \qquad \text{ok}$$

note

1. We showed that $|p - x_n| \leq k|p - x_{n-1}|$; this is called linear convergence and k is called the asymptotic error constant.

2. When x_0 is sufficiently close to p, we can choose $k = |g'(p)|$.

recall : $f(x) = x^2 - 3$, $p = \sqrt{3} = 1.73205$

$$g_1(x) = \frac{3}{x} \Rightarrow g_1'(x) = -\frac{3}{x^2} \Rightarrow |g_1'(p)| = 1 \;:\; \text{diverges}$$

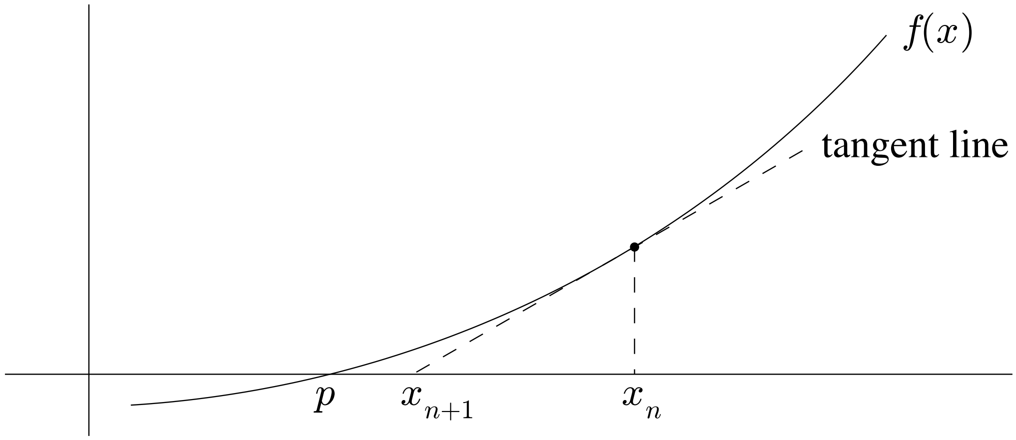
$$g_2(x) = x - (x^2 - 3) \Rightarrow g_2'(x) = 1 - 2x \Rightarrow |g_2'(p)| = 2.4641 \;:\; \text{diverges}$$

$$g_3(x) = x - \left(\frac{x^2 - 3}{2}\right) \Rightarrow g_3'(x) = 1 - x \Rightarrow |g_3'(p)| = 0.73205 \;:\; \text{converges}$$

3. The bisection method also converges linearly, with $k = \frac{1}{2}$.

section 2.4 Newton's method

idea : local linear approximation



$$\text{slope} = f'(x_n) = \frac{0 - f(x_n)}{x_{n+1} - x_n} \;\Rightarrow\; x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

ex : $f(x) = x^2 - 3 \;\Rightarrow\; x_{n+1} = x_n - \dfrac{x_n^2 - 3}{2x_n}$

n	x_n	$f(x_n)$	$\lvert p - x_n \rvert$
0	1.5	-0.75	0.23205081
1	1.75	0.0625	0.01794919
2	1.73214286	0.00031888	0.00009205
3	1.73205081	0.00000001	0.00000001

note

Newton's method is an example of fixed point iteration, $x_{n+1} = g(x_n)$, where the iteration function is $g(x) = x - \dfrac{f(x)}{f'(x)}$.

Then $g'(x) = 1 - \dfrac{f'(x)^2 - f(x)\cdot f''(x)}{f'(x)^2} \Rightarrow g'(p) = 1 - \dfrac{f'(p)^2 - f(p)\cdot f''(p)}{f'(p)^2} = 0.$

Here we assumed that $f(p) = 0, f'(p) \neq 0$, i.e. p is a simple root of $f(x)$. (This is the most common case). This implies that Newton's method converges faster than linearly; in fact we have $|p - x_{n+1}| \leq C|p - x_n|^2$, i.e. quadratic convergence.

pf

$$p - x_{n+1} = g(p) - g(x_n) = \cancel{g(p)} - \big(\cancel{g(p)} + \cancel{g'(p)}(x_n - p) + O(x_n - p)^2\big) \qquad \text{ok}$$

<u>ex</u> : page 102, volume of chlorine gas

P : pressure , V : volume , T : temperature

$PV = nRT$: ideal gas law

n : number of moles present

R : universal gas constant , $R = 0.08206$ atm · liter/(mole · K)

$\left(P + \frac{n^2a}{V^2}\right)(V - nb) = nRT$: van der Waals equation

a : accounts for intermolecular attractive forces , $a = 6.29$ atm · liter^2/mole^2

b : accounts for intrinsic volume of gas molecules , $b = 0.0562$ liter/mole

Take $n = 1$ mole, $P = 2$ atm, $T = 313$ K, and find V by Newton's method with starting guess V_0 given by the ideal gas law.

$$f(V) = \left(P + \frac{n^2a}{V^2}\right)(V - nb) - nRT \ , \ f'(V) = \left(P + \frac{n^2a}{V^2}\right) + \left(\frac{-2n^2a}{V^3}\right)(V - nb)$$

n	V_n	
0	12.84238999999998	
1	12.651154813406302	
2	12.651099337119016	: slightly less than V_0 given by ideal gas law

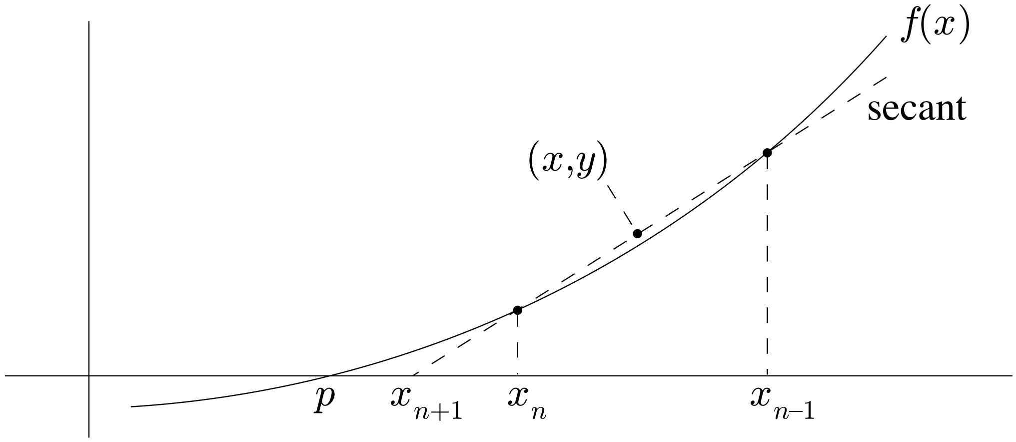
We see that V_0 has 2 correct digits and V_1 has 5 correct digits. How many correct digits does V_2 have? (hw)

<u>note</u>

1. alternative derivation : $\cancelto{0}{f(x_{n+1})} = f(x_n) + f'(x_n)(x_{n+1} - x_n) + \cancelto{0}{\cdots\cdots}$

2. Newton's method converges rapidly, but it requires extra work to compute $f'(x_n)$. Is there an alternative?

<u>section 2.5</u> secant method



slope : $\dfrac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}$, equation : $\dfrac{y-f(x_n)}{x-x_n}=\dfrac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}$

$$(x,y)=(x_{n+1},0) \;\Rightarrow\; x_{n+1}=x_n-\frac{f(x_n)}{\left(\dfrac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}\right)} \quad : \text{ secant method}$$

note

1. The secant method requires two starting values, x_0, x_1.

2. It can be shown that $|p-x_n| \le C|p-x_{n-1}|^{1.6}$, so the secant method converges faster than fixed-point iteration, but slower than Newton's method.

summary

method	rate of convergence	cost per step
bisection	linear , $k=\frac{1}{2}$	$f(x_n)$
fixed-point iteration	linear , $k=\|g'(p)\|$	$g(x_n)$
Newton	quadratic	$f(x_n), f'(x_n)$
secant	between linear and quadratic	$f(x_n)$

note : Bisection is guaranteed to converge if the initial interval contains a root, but the other methods can be very sensitive to the choice of x_0.

rootfinding for nonlinear systems

ex : page 141, chemical reactions

$$\left.\begin{array}{r} 2A+B \rightleftharpoons C \\ A+D \rightleftharpoons C \end{array}\right\} \; : \text{ reversible reactions for reactants } A, B, D \text{ and product } C$$

a_0, b_0, d_0 : initial concentrations (moles/liter) in chemical reactor (known)

c_1, c_2 : equilibrium concentrations of C produced by each reaction (unknown)

k_1, k_2 : equilibrium reaction constants (known)

These variables are related by the Law of Mass Action.

compound	equilibrium concentration
A	$a_0-2c_1-c_2$
B	b_0-c_1
C	c_1+c_2
D	d_0-c_2

$$\Rightarrow \left\{\begin{array}{l} k_1=\dfrac{c_1+c_2}{(a_0-2c_1-c_2)^2(b_0-c_1)} \\ k_2=\dfrac{c_1+c_2}{(a_0-2c_1-c_2)(d_0-c_2)} \end{array}\right.$$

Hence to find c_1, c_2 we need to solve a system of nonlinear equations.

<u>Newton's method for nonlinear systems</u>

$f(x,y)=0\ ,\ \ g(x,y)=0$

Given (x_n,y_n), we want to find (x_{n+1},y_{n+1}).

$$\cancelto{0}{f(x_{n+1},y_{n+1})} \;=\; f(x_n,y_n) + \frac{\partial f}{\partial x}(x_n,y_n)(x_{n+1}-x_n) + \frac{\partial f}{\partial y}(x_n,y_n)(y_{n+1}-y_n) + \cancelto{0}{\cdots\cdots\cdots}$$

$$\cancelto{0}{g(x_{n+1},y_{n+1})} \;=\; g(x_n,y_n) + \frac{\partial g}{\partial x}(x_n,y_n)(x_{n+1}-x_n) + \frac{\partial g}{\partial y}(x_n,y_n)(y_{n+1}-y_n) + \cancelto{0}{\cdots\cdots\cdots}$$

$$\Rightarrow \quad \underbrace{\begin{pmatrix} f_x & f_y \\ g_x & g_y \end{pmatrix}\Bigg|_{(x_n,y_n)}}_{\uparrow\ \text{Jacobian matrix}} \cdot \begin{pmatrix} x_{n+1}-x_n \\ y_{n+1}-y_n \end{pmatrix} = \begin{pmatrix} -f(x_n,y_n) \\ -g(x_n,y_n) \end{pmatrix}$$

This equation has the form $Ax=b$, where A is a given matrix, b is a given vector, and we must solve for the vector x.