Calculus IB: Lecture 17

Luo Luo

Department of Mathematics, HKUST

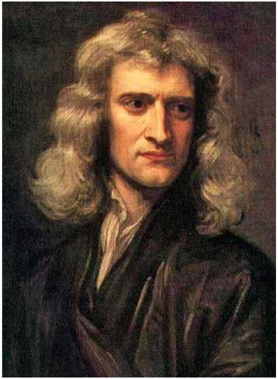
http://luoluo.people.ust.hk/

Outline

Outline

Newton's Method

Newton's method, also known as the Newton–Raphson method, named after Isaac Newton and Joseph Raphson (who can provide a picture/photo of Raphson?), is a root-finding algorithm.



Sir Isaac Newton (1642-1726), the greatest scientist of all time.

Newton's Method

Newton's method is a simple usage of the tangent lines in finding approximate solutions of a non-linear equation

$$f(x) = 0,$$

where f is differentiable and defined on real numbers.

Note that non-linear equation may have no closed form solution, e.g.

$$f(x) = x^5 - x + 1 = 0.$$

In many optimization problems, one important step is finding critical point, which just corresponds to solving a non-linear equation without closed form solution.

Newton's Method

The Newton's method generates sequence

$$x_0, x_1, x_2, x_3 \dots$$

such that $f(x_k)$ converges to 0 with increasing k.

The basic idea is applying linear approximation on $f(x)$ at given x_k

$$f(x) \approx f(x_k) + f'(x_k)(x - x_k).$$

Then we solve the linear equation

$$f(x_k) + f'(x_k)(x - x_k) = 0$$

is an approximation of solving $f(x) = 0$.

Newton's Method

Let x_{k+1} be the solution of

$$f(x_k) + f'(x_k)(x - x_k) = 0.$$

Note that $y = g_k(x) = f(x_k) + f'(x_k)(x - x_k)$ is a linear function whose slope is $f'(x_k)$ and graph passes the point $(x_k, f(x_k))$

Suppose $f'(x_k) \neq 0$, then we have

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}. \tag{1}$$

Newton's method iterates (1) from a suitable initial point x_0.

Roughly speaking, if initial point x_0 is not too far away from a root of $f(x) = 0$, Newton's method produces a sequence $x_1, x_2, x_3, \ldots,$ which may get closer and closer to an exact root of $f(x) = 0$.

Example

Find an approximate positive root of $x^3 - 2 = 0$ by the Newton's method. (exact root $\sqrt[3]{2} = 1.25992104989\cdots$ which can be obtained by calculator)

Let $f(x) = x^3 - 2$, and hence $f'(x) = 3x^2$ and the iteration formula is

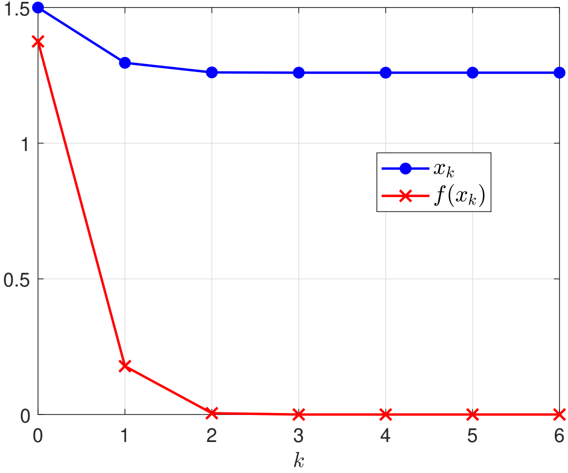
$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^3 - 2}{3x_k^2}$$

k	x_k	$f(x_k) = x_k^3 - 2$	$f'(x_k) = 3x_k^2$	$x_{k+1} = x_k - \frac{x_k^3-2}{3x_k^2}$
0	1.500000000	1.375000000	6.750000000	1.296296296
1	1.296296296	0.178275669	5.041152263	1.260932225
2	1.260932225	0.004819286	4.769850226	1.259921861
3	1.259921861	0.000003861	4.762209284	1.259921050
4	1.259921050	0.000000000	4.762203156	1.259921050
5	1.259921050	0.000000000	4.762203156	1.259921050
6	1.259921050	0.000000000	4.762203156	1.259921050

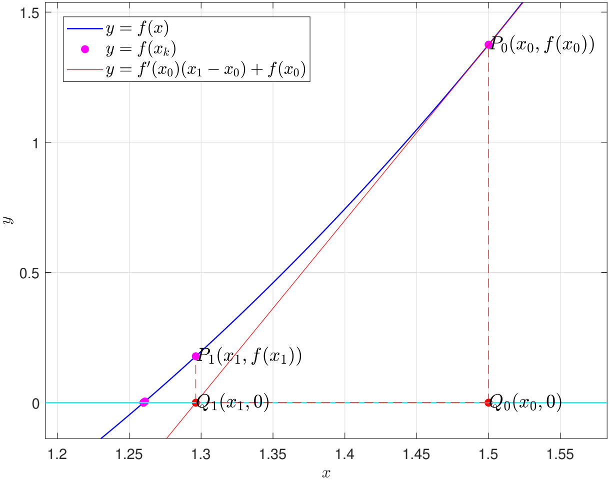
An approximate value of $\sqrt[3]{2}$ is 1.259921050.

Newton's Method

Convergence behavior of x_k and $f(x_k)$ with iterations:



Newton's Method



$$x_0 - x_1 = Q_1 Q_0 = \frac{P_0 Q_0}{\tan \angle P_0 Q_1 Q_0} = \frac{f(x_0)}{f'(x_0)} \Longrightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Newton's Method

Exercise

Find an approximate value of the root of the equation

$$\cos x - x = 0$$

by the Newton's Method (using calculator or MATLAB).

Outline

Convergence Analysis

Above example show that Newton's method

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

produces a sequence

$$x_1, x_2, x_3, \ldots$$

which converge to the solution x^* such that $f(x^*) = 0$.

We impose the following assumptions to further analysis

1. the function f is differentiable
2. there exists positive μ such that $|f'(x)| \geq \mu$ for all x
3. there exists positive C such that $|f''(x)| \leq C$ for all x
4. the initial point is NOT far away form x^*

(We typically suppose $f'(x)$ is C-Lipschitz continuous, rather than $|f''(x)| \leq C$, but the related analysis needs some techniques beyond MATH 1013.)

Convergence Analysis

Mean Value Theorem

If f is twice differentiable on (a, b) and continuous on $[a, b]$, then

$$f(b) - f(a) - f'(a)(b - a) = \frac{f''(c)}{2}(b - a)^2$$

for some $c \in (a, b)$. If $a > b$ and function f is twice differentiable on (b, a) and continuous on $[b, a]$, we can also find c between a and b satisfies above inequality.

The mean value theorem means there exists c_k between x_k and x^* that

$$f(x^*) - f(x_k) - f'(x_k)(x^* - x_k) = \frac{f''(c_k)}{2}(x^* - x_k)^2.$$

Using the iteration of Newton's Method

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

and the result from mean value theorem

$$f(x^*) - f(x_k) - f'(x_k)(x^* - x_k) = \frac{f''(c_k)}{2}(x^* - x_k)^2,$$

we have

$$\begin{aligned}
|x_{k+1} - x^*| &= \left| x_k - \frac{f(x_k)}{f'(x_k)} - x^* \right| \\
&= \frac{1}{|f'(x_k)|} \left| f'(x_k) \cdot (x_k - x^*) - (f(x_k) - f(x^*)) \right| \\
&= \frac{1}{|f'(x_k)|} \left| \frac{f''(c_k)}{2}(x^* - x_k)^2 \right| = \frac{|f''(c_k)|}{2|f'(x_k)|} |x_k - x^*|^2
\end{aligned}$$

Convergence Analysis

Since $|f'(x)| \geq \mu$ and $|f''(x)| \leq C$, we have

$$|x_{k+1} - x^*| = \frac{|f''(c_k)|}{2|f'(x_k)|}|x_k - x^*|^2 \leq \frac{C}{2\mu}|x_k - x^*|^2.$$

Then we have (try to prove the last one by induction)

$$\begin{cases} |x_1 - x^*| \leq \dfrac{C}{2\mu}|x_0 - x^*|^2 \\ |x_2 - x^*| \leq \dfrac{C}{2\mu}|x_1 - x^*|^2 \leq \left(\dfrac{C}{2\mu}\right)^3 |x_0 - x^*|^4 \\ |x_3 - x^*| \leq \dfrac{C}{2\mu}|x_2 - x^*|^2 \leq \left(\dfrac{C}{2\mu}\right)^7 |x_0 - x^*|^8 \\ \qquad \cdots\cdots \\ |x_k - x^*| \leq \dfrac{C}{2\mu}|x_{k-1} - x^*| \leq \left(\dfrac{C}{2\mu}\right)^{2^k - 1} |x_0 - x^*|^{2^k} \end{cases}$$

Convergence Analysis

In summary Newton's method

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

from x_0 holds that

$$|x_k - x^*| \leq \left(\frac{C}{2\mu}\right)^{2^k-1} |x_0 - x^*|^{2^k} \leq \frac{2\mu}{C}\left(\frac{C}{2\mu}|x_0 - x^*|\right)^{2^k}.$$

If $\dfrac{C}{2\mu}|x_0 - x^*| < 1$, the sequence will $x_1, \dots, x_k$ converges to x^* very fast.

Otherwise, it may diverge. Hence, we should start with x_0 such that

$$|x_0 - x^*| < \frac{2\mu}{C}.$$

Outline

Newton's Method in Convex Optimization

The basic idea is using Newton's method to solve $f'(x) = 0$, since the equation has no closed form solution in many situations.

Theorem

If function f is convex and differentiable over an interval I. Then any point x^* that satisfies $f'(x^*) = 0$ holds that $f(x^*)$ is a global minimum.

We consider the simple case that $I = (-\infty, \infty)$.

Newton's Method in Convex Optimization

Impose previous assumption on f':

1. the function f' differentiable
2. there exists positive μ such that $|f''(x)| \geq \mu$ for all x
3. there exists positive C such that $|f'''(x)| \leq C$ for all x
4. the initial point x_0 satisfies $|x_0 - x^*| < \frac{2\mu}{C}$

In the view of convex optimization:

1. the function f is twice differentiable
2. the function f is strongly convex with factor $\mu > 0$, i.e. $f''(x) \geq \mu$.
3. there exists positive C such that $|f'''(x)| \leq C$ for all x
4. the initial point x_0 satisfies $|x_0 - x^*| < \frac{2\mu}{C}$

The third one typically can be replaced by "there exists $C > 0$ such that for any x_1 and x_2, we have $|f''(x_1) - f''(x_2)| \leq C|x_1 - x_2|$".

According to the analysis of Newton's method, we have

$$|x_k - x^*| \leq \frac{2\mu}{C}\left(\frac{C}{2\mu}|x_0 - x^*|\right)^{2^k}.$$

By assuming $f''(x) \leq L$, we can establish the result of function value

$$f(x_k) - f(x^*) - f'(x^*)(x_k - x^*) = \frac{f''(c_k)}{2}(x^* - x_k)^2$$

$$\Longrightarrow f(x_k) - f(x^*) \leq \frac{L}{2}(x^* - x_k)^2 \leq \frac{L}{2}\left(\frac{2\mu}{C}\right)^2\left(\frac{C}{2\mu}|x_0 - x^*|\right)^{2^{k+1}}$$

$$\Longrightarrow f(x_k) - f(x^*) \leq \frac{2\mu^2 L}{C^2}\left(\frac{C}{2\mu}|x_0 - x^*|\right)^{2^{k+1}}$$

Newton's Method in Convex Optimization

According to the analysis of Newton's method, we have

$$f(x_k) - f(x^*) \leq \frac{2\mu^2 L}{C^2} \left(\frac{C}{2\mu} |x_0 - x^*| \right)^{2^{k+1}}.$$

If we desire $f(x_k) \approx f(x^*)$ such that $f(x_k) - f(x^*) \leq \varepsilon$, which requires

$$\begin{aligned}
& \frac{2\mu^2 L}{C^2} \left(\frac{C}{2\mu} |x_0 - x^*| \right)^{2^{k+1}} \leq \varepsilon \\
\Longrightarrow & \left(\frac{C}{2\mu} |x_0 - x^*| \right)^{2^{k+1}} \leq \frac{C^2 \varepsilon}{2\mu^2 L} \\
\Longrightarrow & 2^{k+1} \geq \log_{\frac{C}{2\mu}|x_0 - x^*|} \left(\frac{C^2 \varepsilon}{2\mu^2 L} \right) \\
\Longrightarrow & k \geq \log_2 \left(\log_{\frac{C}{2\mu}|x_0 - x^*|} \left(\frac{C^2 \varepsilon}{2\mu^2 L} \right) \right) - 1.
\end{aligned}$$

Strengths of Newton's Method

Note that

$$k \geq \log_2 \left(\log_{\frac{C}{2\mu}|x_0 - x^*|} \left(\frac{C^2 \varepsilon}{2\mu^2 L} \right) \right) - 1.$$

means we only need very few even if ε is very small.

Since $\frac{C}{2\mu}|x_0 - x^*| < 1$, we suppose $\frac{C}{2\mu}|x_0 - x^*| = 0.999 \approx 1$.

We also assume ε is very small such that $\frac{C^2 \varepsilon}{2\mu^2 L} = 10^{-8}$.

Then we only require $k \geq 13.1683$.

Strengths of Newton's Method

Recall that gradient descent holds that

$$f(x_k) - f(x^*) \leq \left(1 - \frac{\mu}{L}\right)^k (f(x_0) - f(x^*))$$

for strongly convex and differentiable f.

To find x_k such that $f(x_k) - f(x^*) \leq \varepsilon$, it needs

$$\begin{aligned} &\left(1 - \frac{\mu}{L}\right)^k (f(x_0) - f(x^*)) \leq \varepsilon \\ \Longrightarrow &\left(1 - \frac{\mu}{L}\right)^k \leq \frac{f(x_0) - f(x^*)}{\varepsilon} \\ \Longrightarrow & k \geq \log_{\left(1-\frac{\mu}{L}\right)} \left(\frac{f(x_0) - f(x^*)}{\varepsilon}\right) \end{aligned}$$

Strengths of Newton's Method

If it is desired a very very accuracy approximation, we only interested in how ε affects k since ε is much smaller than other terms.

Then gradient descent needs

$$k \geq \log_{C_1}(C_2\varepsilon)$$

and Newton's method needs

$$k \geq \log_2\left(\log_{C_3}(C_4\varepsilon)\right).$$

We have (try to show that by L'Hôpital's rule as an exercise)

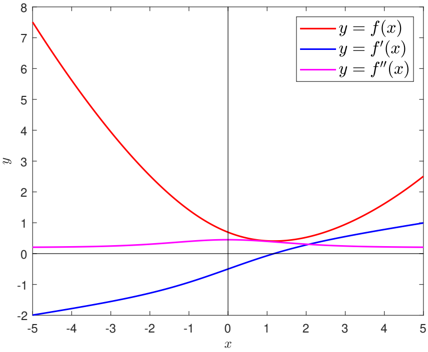
$$\lim_{\varepsilon\to0^+} \frac{\log_2\left(\log_{C_3}(C_4\varepsilon)\right)}{\log_{C_1}(C_2\varepsilon)} = 0.$$

Strengths of Newton's Method

Consider the example of finding minimum of

$$f(x) = \frac{x^2}{10} + \ln(1 + e^{-x}),$$

we have $f'(x) = \dfrac{x}{5} - \dfrac{e^{-x}}{e^{-x} + 1}$ and $f''(x) = \dfrac{e^{2x} + 7e^2 + 1}{5(e^x + 1)^2} \leq \dfrac{9}{2}$.



Strengths of Newton's Method

We run gradient descent

$$x_{k+1} = x_k - \frac{1}{L} f'(x_k)$$

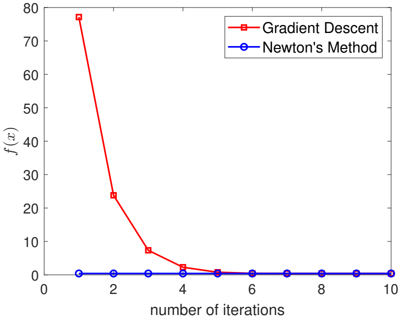
with $L = \frac{9}{2}$ and Newton's method

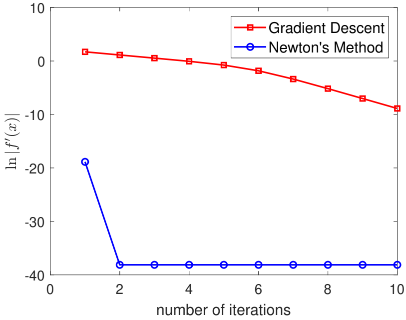
$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}.$$

We select $x_0 = 10$ and run 10 iterations for both algorithms.

Strengths of Newton's Method

The convergence behavior of $f(x)$ and $\ln|f'(x)|$. In theoretical, $|f'(x)|$ should tend to 0 which leads to $\ln|f'(x)| \to -\infty$, but the computer cannot present too small magnitude of a real number if it is not 0.





Weakness of Newton's Method

The initial point x_0 of Newton's method should be near x^*:

$$\frac{C}{2\mu}|x_0 - x^*| < 1.$$

Unfortunately, there is no good strategy to select x_0 for general f since we do not know what is x^* at first.

In other words, the convergence of Newton's method is local, not global.

On the other hand, gradient descent could converge to the optimal solution with any initial point x_0 (global convergence).

Weakness of Newton's Method

Newton's method depends on the twice differentiability while gradient descent only requires first differentiability.

Newton's method works only if $f''(x_k) \neq 0$ which is unnecessary for gradient descent.

If f has many input variables, the computational cost of Newton's method is much more expensive than gradient descent.

Weakness of Newton's Method

Consider using Newton's method to solve

$$f(x) = x^2 - 5 = 0.$$

If we let $x_0 = 0$, then $f'(x_0) = 2x_0 = 0$ and the update

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \quad \text{is undefined.}$$

Weakness of Newton's Method

Newton's method may fail even if $f'(x_k) \neq 0$ for any k ~~when x_0 is far way from the solution~~ and x_0 is very close to x^*. Note that that f' may be undefined at x^*.

Exercise

Consider solving

$$f(x) = x^{\frac{1}{3}} = 0$$

by Newton's method with initial point $x_0 = 1$.

Exercise

Compare the convergence of Newton's method with bisection method (Lecture 08). Which one is faster?