

Lecture A1. Math Review

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr



서울과학기술대학교 데이터사이언스학과

1 I. Differentiation and Integration

2 II. Numerical Methods

3 III. Matrix Algebra

4 IV. Series and Others

I. Differentiation and Integration

Differentiation

Definition 1 (differentiation)

Differentiation is the action of computing a derivative.

Definition 2 (derivative)

The derivative of a function $y = f(x)$ of a variable x is a measure of the rate at which the value y of the function changes with respect to (wrt., hereafter) the change of the variable x . It is notated as $f'(x)$ and called derivative of f wrt. x .

Remark 1

If x and y are real numbers, and if the graph of f is plotted against x , the derivative is the slope of this graph at each point.

Definition 3 (differentiable)

If $\lim_{h \rightarrow 0} \frac{f(x+h/2)-f(x-h/2)}{h}$ exists for a function f at x , we say the function f is *differentiable at x* . That is, $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h/2)-f(x-h/2)}{h}$. If f is differentiable for all x , then we say f is *differentiable (everywhere)*.

Remark 2

The followings are popular derivatives.

- $f(x) = x^p \Rightarrow f'(x) = px^{p-1}$ (polyomial)
- $f(x) = e^x \Rightarrow f'(x) = e^x$ (exponential)
- $f(x) = \log(x) \Rightarrow f'(x) = 1/x$ (log function; not differentiable at $x = 0$)

Theorem 1

Differentiation is linear. That is, $h(x) = f(x) + g(x)$ implies $h'(x) = f'(x) + g'(x)$.

Theorem 2 (differentiation of product)

If $h(x) = f(x)g(x)$, then $h'(x) = f'(x)g(x) + f(x)g'(x)$.

Exercise 1

Suppose $f(x) = xe^x$, find $f'(x)$.

Theorem 3 (differentiation of fraction)

If $h(x) = \frac{f(x)}{g(x)}$, then $h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$.

Theorem 4 (composite function)

If $h(x) = f(g(x))$, then $h'(x) = f'(g(x)) \cdot g'(x)$.

Exercise 2

Suppose $f(x) = e^{2x}$, find $f'(x)$.

Integration

Definition 4 (integration)

Integration is the computation of an integral, which is a reverse operation of differentiation up to an additive constant.

Definition 5 (antiderivative)

Let's say a function f is a derivative of g , or $g'(x) = f(x)$, then we say g is an *antiderivative* of f , written as $g(x) = \int f(x)dx + C$, where C is a integration constant.

Remark 3

The followings are popular antiderivatives.

- For $p \neq -1$, $f(x) = x^p \Rightarrow \int f(x)dx = \frac{1}{p+1}x^{p+1} + C$ (polynomial)
- $f(x) = \frac{1}{x} \Rightarrow \int f(x)dx = \log(x) + C$ (fraction)
- $f(x) = e^x \Rightarrow \int f(x)dx = e^x + C$ (exponential)
- $f(x) = \frac{g'(x)}{g(x)} \Rightarrow \int f(x)dx = \log(g(x)) + C$ (See Theorem 4 above)

Exercise 3

Derive $\int f'(x)g(x) dx = f(x)g(x) - \int f(x)g(x)' dx$. (Hint: Use Theorem 2 above.)

Exercise 4

Find $\int x e^x dx$, and evaluate $\int_0^1 x e^x dx$. (Hint: Use Exercise 3 above.)

II. Numerical Methods

Differentiation

- Oftentimes, finding analytic derivative is hard, but finding numerical derivative is often possible.

Definition 6

For a function f and small constant h ,

- $f'(x) \approx \frac{f(x+h)-f(x)}{h}$ (*forward difference formula*)
- $f'(x) \approx \frac{f(x)-f(x-h)}{h}$ (*backward difference formula*)
- $f'(x) \approx \frac{f(x+h)-f(x-h)}{2h}$ (*centered difference formula*)

Solving an equation

- For the rest of this section, we consider a nonlinear and differentiable (thus, continuous) function $f : \mathbb{R} \rightarrow \mathbb{R}$, we aim to find a point $x^* \in \mathbb{R}$ such that $f(x^*) = 0$. We call such x^* as a *solution* or a *root*.

Bisection Method

- The *bisection* method aims to find a very short interval $[a, b]$ in which f changes a sign.
- Why? Changing a sign from a to b means the function crosses the $\{y = 0\}$ -axis, (a.k.a. x -axis), at least once. It means x^* such that $f(x^*) = 0$ is in this interval. Since $[a, b]$ is a very short interval, We may simply say $x^* = \frac{a+b}{2}$.

Definition 7 (sign function)

$\text{sgn}(\cdot)$ is called a *sign function* that returns 1 if the input is positive, -1 if negative, and 0 if zero.

Bisection algorithm

- Let tol be the maximum allowable length of the *short interval* and an initial interval $[a, b]$ be such that $sgn(f(a)) \neq sgn(f(b))$.
- The *bisection algorithm* is the following.

```
1: while  $((b - a) > tol)$  do
2:    $m = \frac{a+b}{2}$ 
3:   if  $sgn(f(a)) = sgn(f(m))$  then
4:      $a = m$ 
5:   else
6:      $b = m$ 
7:   end
8: end
```

- At each *iteration*, the interval length is halved. As soon as the interval length becomes smaller than tol , then the algorithm stops.

Newton Method

- The bisection technique makes no use of the function values other than their signs, resulting in slow but sure convergence.
- More rapid convergence can be achieved by using the function values to obtain a more accurate approximation to the solution *at each iteration*.
- Newton method is a method that uses both the function value and derivative value.

- Newton method approximates the function f near x_k by the tangent line at $f(x_k)$.

1: $x_0 =$ initial guess

2: for $k=0,1,2,\dots$

3: $x_{k+1} = x_k - f(x_k)/f'(x_k)$

4: break if $|x_{k+1} - x_k| < tol$

5: end

- Root-finding numerical methods such as bisection method and newton method has a few common properties.
 - ① It is characterized as a *iterative process* (such as $x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow \dots$).
 - ② In each *iteration*, the current candidate *gets closer* to the true value.
 - ③ It converges. That is, it is theoretically reach the *exact value* up to tolerance.
- Many iterative numerical methods share the properties above.
- The famous back propagation in deep neural network is also motivated by Newton method.
- Major algorithms for dynamic programming are called *policy iteration* and *value iteration* that also share the properties above.

III. Matrix Algebra

Matrix multiplication

Exercise 5

Solve the followings.

$$(.6 \quad .4) \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix} =$$

Exercise 6

What is P^2 ?

$$P = \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix}$$

Solution to system of linear equations

Exercise 7

Solve the followings.

$$(\pi_1 \quad \pi_2) \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix} = (\pi_1 \quad \pi_2)$$

$$\pi_1 + \pi_2 = 1$$

Exercise 8

Solve the following system of equations.

$$x = y$$

$$y = 0.5z$$

$$z = 0.6 + 0.4x$$

$$x + y + z = 1$$

Exercise 9

Solve the following system of equations.

$$(\pi_0 \quad \pi_1 \quad \pi_2) \begin{pmatrix} -2 & 2 \\ 3 & -5 & 2 \\ & 3 & -3 \end{pmatrix} = (0 \quad 0 \quad 0)$$

$$\pi_0 + \pi_1 + \pi_2 = 1$$

Exercise 10

Solve the following system of equations.

$$(\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4) \begin{pmatrix} .7 & .3 \\ .5 & .5 \\ & .6 & .4 \\ & .3 & .7 \end{pmatrix} = (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4)$$

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

Exercise 11

Solve following and express π_i for $i = 0, 1, 2, \dots$

$$\begin{aligned}
 \pi_0 + \pi_1 + \pi_2 + \dots &= 1 \\
 0.02\pi_0 + 0.02\pi_1 + 0.02\pi_2 + \dots &= \pi_0 \\
 0.98\pi_0 &= \pi_1 \\
 0.98\pi_1 &= \pi_2 \\
 0.98\pi_2 &= \pi_3 \\
 \dots &= \dots
 \end{aligned}$$

IV. Series and Others

Exercise 12 (Infinite geometric series)

Simplify the following. When $|r| < 1$, $S = a + ar + ar^2 + ar^3 + \dots$

Exercise 13 (Finite geometric series)

Simplify the following. When $r \neq 1$, $S = a + ar + ar^2 + ar^3 + \dots + ar^{n-1}$

Exercise 14 (Power series)

Simplify the following. When $|r| < 1$, $S = r + 2r^2 + 3r^3 + 4r^4 + \dots$

Formulation of time varying function

Exercise 15

During the first hour ($0 \leq t \leq 1$), $\lambda(t)$ increases linearly from 0 to 60. After the first hour, $\lambda(t)$ is constant at 60. Draw plot for $\lambda(t)$ and express the function in math form.

"Man can learn nothing unless he proceeds from the known to the unknown. - Claude Bernard"