

## Chapter II.B. The Chain Rule

**Preface:** To find the derivative of  $f(x) = [\sin(x)]^2$  and  $g(x) = \exp(2x) = e^{2x} = [e^x]^2$  you could view these functions as the products,  $\sin(x) \cdot \sin(x)$  or  $e^x \cdot e^x$ . With this view you can apply the product rule to obtain the results,  $f'(x) = 2\sin(x) \cdot \cos(x)$  or  $g'(x) = 2e^{2x}$ . It is a more subtle problem to find the derivative of the functions  $P(x) = \sin(x^2)$  and  $Q(x) = \exp(x^2)$ . These are **not products** but what are sometimes called **the composition of functions**. What this means is that  $P(x)$  and  $Q(x)$  computed by first squaring  $x$  and then finding the exponential or sine function of that result. In this section we'll investigate the rule that allows us to find derivatives for composite functions. Since these functions involve **a chain of operations** to compute their values, this rule is called the **Chain Rule**. Before going any further, here is an example that is typical of situations that call for use of the chain rule.

---

**Example II.B.1.** Suppose my car consumes 1/20 of a gallon of gasoline in traveling one mile ( or, if you prefer, the car goes twenty miles on a gallon of gasoline). If we let  $G$  be the amount of gas consumed in traveling  $s$  miles then we can express this rate of gas consumption by saying

$$dG/ds = 1/20 .$$

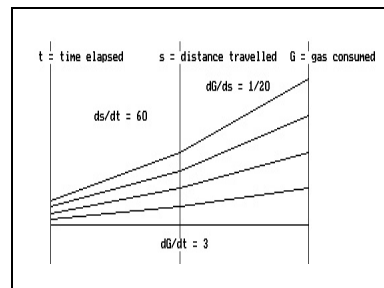
I am driving this same car on the highway at a velocity of 60 miles per hour. Thus if  $s$  denotes the position of the car at time  $t$  we can express this by the equation

$$ds/dt = 60.$$

What I'm interested in knowing is how fast with respect to time  $t$  is the car consuming the gas  $G$ , that is, what is  $dG/dt$ ? It shouldn't take long for you to see why the time rate for consuming the gasoline is 3 gallons per hour. To find  $dG/dt$  we multiplied 1/20 (gal/mi) by 60 (mi/hr), that is, in Leibniz notation:

$$\frac{dG}{dt} = \frac{dG}{ds} \cdot \frac{ds}{dt} .$$

**Comment:** The example uses the variable  $G$  ambiguously, once to name the gas consumed as a function of time, and then to denote the gas consumed as a function of distance. The variable  $s$  is also used in two ways, once as the **controlling variable** for the gas consumption and once as the **variable controlled by time**. [See Figure \*\*\*.] What is more, even though **the derivative is not a quotient but the limit of quotients**, the Leibniz notation (which does suggest a quotient) supports the result from the visual appearance of cancellation of terms in a product of quotients. The



**Figure 1**

way the units seem to cancel  $[\frac{gal}{mi} \cdot \frac{mi}{hr} = \frac{gal}{hr}]$  also supports the apparent correctness of the result.

**Discussion:** One of the key ideas in using the rate interpretation of the derivative is that at any particular point the derivative of a function can be interpreted as the instantaneous rate of change of a controlled (dependent) variable with respect to change of a controlling (independent) variable. This view suggests that what works for constant rates over a long period of time is very likely to hold true

for the derivatives at a single point or instant of time. We will pursue the application of the chain rule to related rates in a later section. We'll carry this view of derivatives as rates forward in the next example to suggest a solution to what the derivative of  $P(t) = \sin(t^2)$  should be.

**Example II.B.2.** Suppose that  $G = \sin(t^2)$ . We can think of  $G$  as a gas consumption situation where the gas consumed depends on time.  $G$  is composed of the sine function and the squaring function. Let's introduce what we will call a **linking variable, called  $s$** , which will play a role to connect the time to the gas consumed. We let  $s = t^2$  so that  $G = \sin(s)$ . Think of  $s$  as the distance traveled. Our problem is to find  $dG/dt$ . The last discussion suggests that we consider this as a rate problem where the rate is constant determined by the component rates also considered as constants. Thus  $ds/dt = 2t$ ,  $dG/ds = \cos(s)$  and hence, as in the last example,

$$\frac{dG}{dt} = \frac{dG}{ds} \cdot \frac{ds}{dt} = \cos(s) \cdot 2t = 2t \cdot \cos(t^2)$$

The answer suggested as correct by this interpretation is justified by

**THEOREM II.4: (The Chain Rule)** If  $g$  is differentiable at  $a$  and  $f$  is differentiable at  $g(a)$  then the function  $P(x) = f(g(x))$  is differentiable at  $a$  and  $P'(a) = f'(g(a)) \cdot g'(a)$ .

In the Leibniz notation the chain rule is usually expressed as follows:

Suppose  $y = f(u)$  and  $u = g(x)$  and  $u$  is differentiable at  $x = a$  while  $y$  is differentiable at  $u = b = g(a)$ . We will refer to the variable  $u$  as the **linking variable in the chain**. Then

$$\left. \frac{dy}{dx} \right|_{x=a} = \left. \frac{dy}{du} \right|_{u=b=f(a)} \cdot \left. \frac{du}{dx} \right|_{x=a}$$

In the operator notation the result is stated sometimes as

$$D_x P(x) = D_u f(b) D_x g(a) = Df(g(a)) \cdot Dg(a)$$

**PSEUDOPROOF:** [The following is a proof except for functions  $g$  where in any interval containing  $a$ ,  $g(x) = g(a)$  for some  $x$ . The difficulty in those cases will be noted later in the discussion. The details of an argument to deal with these exceptions will be found by the interested reader in the appendix to this section.]

Let  $b = g(a)$  and  $k = g(a+h) - g(a) = g(x) - g(a)$  for  $h \neq 0$ .

Note  $g(x) = g(a+h) = g(a) + k = b + k$ . See Figure II.4.

From the assumption that  $g$  is differentiable at  $a$ , we have that  $g$  is also continuous at  $a$ . Thus we can conclude that as  $h \rightarrow 0$ ,  $k \rightarrow 0$ .

We'll follow the usual four steps in finding the derivative of  $P$  at  $a$ :

Step I:  $P(a+h) = f(g(a+h))$

-  $P(a) = f(g(a))$

Step II:  $P(a+h) - P(a) = f(g(a+h)) - f(g(a)) = f(b+k) - f(b)$ .

Now assume that  $k \neq 0$ . [As noted above, this is a major assumption for some functions.]

Then

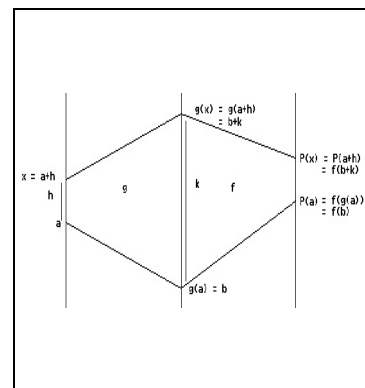


Figure 2

$$P(a+h) - P(a) = \frac{[f(b+k) - f(b)] \cdot k}{k}$$

$$\text{Therefore } P'(a) = \lim_{h \rightarrow 0} \frac{P(a+h) - P(a)}{h}$$

$$= \lim_{\substack{h \rightarrow 0 \\ k \rightarrow 0}} \frac{[f(b+k) - f(b)] \cdot k}{h}$$

$$= \lim_{\substack{h \rightarrow 0 \\ k \rightarrow 0}} \frac{[f(b+k) - f(b)]}{k} \cdot \frac{[g(a+h) - g(a)]}{h}$$

$$= f'(b) \cdot g'(a)$$

$$= f'(g(a)) \cdot g'(a).$$

**EOP.**

Here are some examples of the use of the chain rule, many with some functions that are usually written with the function notation. Here are more examples of the use of the chain rule using trigonometric functions. Recall that  $\sin'(x) = \cos(x)$  and  $\sec'(x) = \sec(x) \tan(x)$  have been justified in chapter I and in the exercises of the last section.

**Example II.B.3.** : Find the derivative for each of the following functions:

(a)  $y = \sin(3x)$  ; (b)  $y = \sin(x^3)$  ; and (c)  $y = [\sin(x)]^3 = \sin^3(x)$ .

**Solution:** (a) Let  $u = 3x$  so that  $y = \sin(u)$ . Then

$$\frac{du}{dx} = 3 \quad \text{and} \quad \frac{dy}{du} = \cos(u),$$

so using the chain rule we have

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \cos(u) \cdot 3 \\ &= 3 \cos(3x). \end{aligned}$$

(b) Let  $u = x^3$  so that  $y = \sin(u)$ . Then

$$\frac{du}{dx} = 3x^2 \quad \text{and} \quad \frac{dy}{du} = \cos(u),$$

so using the chain rule we have

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

$$\begin{aligned}
 &= \cos(u) 3x^2 \\
 &= 3x^2 \cos(x^3).
 \end{aligned}$$

(c) Let  $u = \sin(x)$  so that  $y = u^3$ . Then

$$\frac{du}{dx} = \cos(x) \text{ and } \frac{dy}{du} = 3u^2,$$

so using the chain rule we have

$$\begin{aligned}
 \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\
 &= 3u^2 \cos(x) \\
 &= 3[\sin(x)]^2 \cos(x) = \mathbf{3 \sin^2(x) \cos(x)}.
 \end{aligned}$$

**Example II.B.4.** Find  $dy/dx$  when  $y = \sec(3x^2 + 1)$  and when  $y = (3x^2 + 1)^{1/3}$ .

**Solution:** Let  $u = 3x^2 + 1$ . Then in the first function  $y = \sec(u)$ ,  $du/dx = 6x$  and  $dy/du = \sec(u) \tan(u)$ . Using the chain rule we obtain

$$\begin{aligned}
 \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\
 &= \sec(u) \tan(u) [6x] \\
 &= \mathbf{6x \cdot \sec(3x^2 + 1) \tan(3x^2 + 1)}.
 \end{aligned}$$

For the second function  $y = u^{1/3}$ ,  $du/dx = 6x$  and  $dy/du = 1/3 u^{-2/3}$ . Again, using the chain rule we obtain

$$\begin{aligned}
 \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\
 &= 1/3 u^{-2/3} [6x] \\
 &= \mathbf{6x \cdot 1/3 (3x^2 + 1)^{-2/3} = 2x(3x^2 + 1)^{-2/3}}.
 \end{aligned}$$

**Example II.B.5. :** Find the derivative for each of the following functions: (a)  $y = \exp(3x)$ ; (b)  $y = \exp(x^3)$ ; and (c)  $y = [\exp(x)]^3$ .

**Solution:** (a) Let  $u = 3x$  so that  $y = \exp(u)$ . Then

$$\frac{du}{dx} = 3 \text{ and } \frac{dy}{du} = \exp(u),$$

so using the chain rule we have

$$\begin{aligned}
 \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\
 &= \exp(u) 3 \\
 &= \mathbf{3 \exp(3x)}.
 \end{aligned}$$

(b) Let  $u = x^3$  so that  $y = \exp(u)$ . Then

$$\frac{du}{dx} = 3x^2 \text{ and } \frac{dy}{du} = \exp(u), \text{ so using the chain rule we have}$$

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \exp(u) \cdot 3x^2 \\ &= 3x^2 \exp(x^3). \end{aligned}$$

(c) Let  $u = \exp(x)$  so that  $y = u^3$ . Then

$$\frac{du}{dx} = \exp(x) \text{ and } \frac{dy}{du} = 3u^2,$$

so using the chain rule we have

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= 3u^2 \exp(x) \\ &= 3[\exp(x)]^2 \exp(x) = 3 \exp^3(x). \end{aligned}$$

**Comment :** Two questions that arise in the chain rule's use are

**i) How do I know that I'm supposed to be using the chain rule? and ii) How do I choose the linking variable in a particular problem that is more involved?**

The answer to i) is that you use the chain rule **whenever** you have a complicated expression for a function in which **the last thing computed in evaluating the function is not a sum, difference, product or quotient**. In other words, if none of the arithmetic rules can be applied to find the derivative.

The choice of an expression for the linking variable is also based on the recognition of the function's evaluation. Usually the easiest **choice** for the linking variable **u is the expression of all evaluations done prior to that last computation which suggested the use of the chain rule** in the first place. See if this advice makes sense in the previous and the next examples.

**Example II.B.6.** Suppose  $R(t)$  is a differentiable function with  $R'(t) = 1/(t^2 + 1)$  for all  $t$ . If  $G(t) = R(3t^2 + 1)$  find  $G'(1)$ .

**Solution:** Let  $u(t) = 3t^2 + 1$  so that  $G(t) = R(u(t))$ .

By the chain rule,  $G'(1) = R'(u(1)) \cdot u'(1)$ . Now  $u'(t) = 6t$  so  $u'(1) = 6$ . We also have  $u(1) = 3 + 1 = 4$ , so  $R'(u(1)) = R'(4) = 1/(4^2 + 1) = 1/17$ . So when we put this information together we have  $G'(1) = 1/17 \cdot 6 = 6/17$ .

**Example II.B.7. The derivatives of exponential functions.** Looking back at section I.F we saw there that the derivative of  $e^x$  was itself, i.e.,  $\frac{d(e^x)}{dx} = e^x$ . Although we also showed in that section

that the derivative of any simple exponential function of the form  $b^x$  would have a derivative that was proportional to itself, we only suggested what the proportionality constant is. In particular, let's examine  $f(x) = 2^x$  again to see why that constant is approximately .69315 and then you can proceed



14. a.  $y = \frac{\sin(x^2)}{(x^2+1)^3}$                       b.  $y = \sec(\tan(\sin(x)))$

15. [The reciprocal rule is a special case of the chain rule.] Use the chain rule applied to  $R(x) = 1/Q(x) = [Q(x)]^{-1}$  to justify the reciprocal rule.

You will sometimes find situations where you don't know as much about the function as you do about its derivative. [Many scientific theories are based on understanding how things change, i.e., about the rates at which something changes.] In these situations when the function is involved in a composition the chain rule is very useful in determining a derivative. These ideas will be used more intensively in chapter when you look at models for growth and learning. At present we will use this idea to become more adept at using the chain rule in cases where the functions are expressed with function/argument notation.

In each of the following you may assume that A is a function so that  $A'(x) = \frac{1}{1+x^2}$ .

16. a.  $y = A(5x)$                       b.  $y = A(4x + 5)$   
 17. a.  $y = A(x - 1)$                       b.  $y = A(x^5)$   
 18. a.  $y = [A(x)]^5$                       b.  $y = A(x^2 + 3x + 1)$   
 19. a.  $y = x^3 A(x^2)$                       b.  $y = [A(x^5)]^4$   
 20. a.  $y = [A(x^4)]^5$                       b.  $y = A(5x) [A(x)]^5$

21. Suppose f is differentiable at x. Find the derivatives of the following functions in terms of f(x) and f'(x).

a.  $P_n(x) = [f(x)]^n$                       b.  $G(t) = \sin(f(t))$   
 c.  $R_2(x) = f(f(x))$                       d.  $R_3(x) = f(f(f(x)))$

22. Other interpretations of the chain rule. (Exploration Project)

- a. Sometimes it is convenient to view the derivative  $f'(a)$  as a magnification factor which relates the differences in the values of f for x close to a with the difference between x and a. Use this interpretation to explain the chain rule in terms of composite magnification.
- b. Sometimes it is convenient to view the derivative as related to the circumferences of wheels connected by a chain. For every full turn of the first wheel, the second axle turns a factor K of its circumference. Show that K is the ratio of the radius of the first wheel to the radius of the second wheel. Use this interpretation to explain the chain rule in terms of three wheels connected by two chains.

23. Find all x where  $f(x) = 0$  for each of the following functions:

a.  $f(x) = (x^2 - 4x + 3)^{10}$                       b.  $f(x) = (x^2 + 4x + 5)^{10}$   
 c.  $f(x) = \sin(x^2)$                       d.  $f(x) = \sec(x^2)$

24. Find all  $x$  where  $f'(x) = 0$  for each of the following functions:
- $f(x) = \sqrt{x^2 - 4x + 3}$
  - $f(x) = x \cdot \sqrt{x^2 - 4x + 3}$
  - $f(x) = x \cdot e^{-x}$
  - $f(x) = x \cdot e^{-x^2}$
25. **Rational exponents and the chain rule.** Suppose that  $u = f(x)$  is a differentiable function at  $x$  and  $y = u^5 = x^3$  so  $u = x^{3/5}$ . Then according to the chain rule  $dy/dx = 5u^4 du/dx$  and of course  $dy/dx = 3x^2$  so  $5u^4 du/dx = 3x^2$ . Use this to show that  $du/dx = 3/5 x^{-2/5}$ . Generalize this example to give an argument using the chain rule why the power rule should be true for rational powers of  $x$ .
26. **Slope of the tangent line and the chain rule. Shifts.** The chain rule for some functions makes sense in terms of shifting graphs.
- Draw a graph of  $Y = X^2$  and  $Y = (X - 3)^2$  with the same axes. Draw a line tangent to  $Y = X^2$  at the point  $(3, 9)$  and a second line tangent to  $Y = (X - 3)^2$  at  $(6, 9)$ . Are these lines parallel? Explain your response both in terms of the derivative and in terms of the geometric relation of the two graphs.
  - Using the geometry** of your figures in part a), explain why the slope of a line tangent to the graph of  $Y = X^2$  at  $(a, a^2)$  will be the same as the slope of a line tangent to the graph of  $Y = (X - 3)^2$  at  $(b, (b - 3)^2)$  where  $b = a + 3$  and thus  $D((x - 3)^2) = 2(x - 3)$ .
27. Reread problem 26. Draw figures for  $Y = f(X)$  and  $Y = f(X - 3)$  and explain using geometry why if  $P(x) = f(x - 3)$  then  $P'(x) = f'(x - 3)$ . Generalize your argument for  $P(x) = f(x - a)$ .
28. Suppose  $f(x)$  is differentiable for all  $x$ . Using the definition of the derivative, show that if  $P(x) = f(\beta x)$  where  $\beta \neq 0$  then  $P'(x) = \beta f'(\beta x)$ .
29. Suppose the following charts give the values of  $f, f', g,$  and  $g'$ . Find the following values as indicated.

$x$	$f(x)$	$f'(x)$	$g(x)$	$g'(x)$
0	3	2	2	0
1	2	0	1	3
2	1	1	0	2
3	0	3	3	1

- $Q(x) = f(x)/g(x)$  ; find  $Q'(0)$  and  $Q'(1)$ .
  - $P(x) = f(g(x))$  ; find  $P'(0)$  and  $P'(1)$ .
  - $R(x) = g(f(x))$  ; find  $R'(0)$  and  $R'(1)$ .
  - $S(x) = g(f(g(x)))$  ; find  $S'(1)$ .
30. Suppose  $f$  and  $g$  are differentiable functions with graphs as in Figure \*\*\*. Based on these graphs, estimate the following values as indicated.
- $Q(x) = f(x)/g(x)$ ;  $Q'(0)$  and  $Q'(1)$ .



Figure 3



- b.  $P(x) = f(g(x));$   $P'(0)$  and  $P'(1)$ .
- c.  $R(x) = g(f(x));$   $R'(0)$  and  $R'(1)$ .
- d.  $S(x) = g(f(g(x)));$   $S'(1)$ .

Problems 31 - \*\*\* illustrate how the chain rule can be used algebraically to find out about derivatives of functions described by their relationship to other functions.

31. Suppose  $P$  is a differentiable function where  $P'(x) = P(x)$  for all  $x$ . [Notice that the simple exponential function with base  $e$  satisfies this condition.]
- a. Let  $C(x) = P(g(x))$ . Show that  $C'(x) = g'(x) C(x)$ .
  - b.  $L$  is also a differentiable function with domain all positive real numbers where  $P(L(x))=x$  for all  $x>0$ . Prove, using the chain rule, that  $L'(x) = 1/x$ . What does this result for the function  $L$  suggest about the derivative of the logarithmic function with base  $e$ ?
  - c. Suppose  $P$  is a differentiable function where  $P'(x) = \alpha P(x)$  for all  $x$  and  $\alpha$  is a constant. [Notice that simple exponential functions satisfy this condition.] Now suppose  $L_\alpha$  is also a differentiable function with domain all positive real numbers where  $P(L_\alpha(x))=x$  for all  $x>0$ . Prove, Use the chain rule to show that  $L_\alpha'(x) = 1/(\alpha x)$ .
32. Suppose  $L$  is a differentiable function where  $L'(x) = 1/x$  for all  $x>0$ . [Notice that the natural logarithm function satisfies this condition.]
- a. Let  $M(x) = L(\alpha x)$ . Show that  $M'(x) = 1/x$ .
  - b. [A converse to problem 31a.] Suppose that  $P$  is also a differentiable function with  $P(x) > 0$  for all  $x$ , and that  $L(P(x))=x$  for all  $x$ . Prove  $P'(x)=P(x)$  for all  $x$ .
  - c. Suppose  $f$  and  $g$  are both differentiable functions with  $f(g(x))=x$  for all  $x$ . Prove that  $g'(x) = 1/f'(g(x))$ . Interpret the result in terms of a transformation figure and in terms of the slopes of lines tangent to curves symmetric with respect to the line  $Y=X$ .
33. Suppose  $p(x) = f(g(x))$  and  $p'(1) = 0$ . Which of the following must be true:
- a.  $g'(1) = 0$
  - b.  $g(1) = 0$
  - c.  $f'(g(1))= 0$
  - d.  $f(1) = 0$
  - e.  $f(1) = 0$
  - f. a or d
  - g. a or c
  - h. a and c.

## Appendix II.B. The Proof of the chain rule.

The pseudoproof of the chain rule given in this section can be made into an actual proof for the chain rule by giving an argument for the cases excluded in the pseudoproof. Returning to the pseudoproof, recall that we had  $k = g(a+h)-g(a)$  for  $h \neq 0$ . The difficulties of the pseudoproof arose from the possibility that  $k$  might equal 0 for values of  $h$  arbitrarily close to 0. We'll consider these functions now and see that the chain rule formula still will be true.

First, since we assume that  $g$  is differentiable we know that  $\lim_{h \rightarrow 0} \frac{g(a+h)-g(a)}{h}$  must exist. Our assumption that  $k = 0$  for  $h$

arbitrarily close to 0 means that for these values of  $h$ ,  

$$\frac{g(a+h)-g(a)}{h} = 0. \quad [\text{See Figure II.6.}]$$

Thus  $g'(a) = \lim_{h \rightarrow 0} \frac{g(a+h)-g(a)}{h} = 0$ . [0 is the only possible limit.]

To complete the argument we need only show that  $P'(a) = 0$  also. But for precisely the same  $h$  values that had  $k = g(a+h)-g(a) = 0$ , we have  $g(a+h) = g(a)$ . Thus for these values of  $h$

$$\begin{aligned} P(a+h) - P(a) &= f(g(a+h)) - f(g(a)) \\ &= f(b + k) - f(b) = f(b) - f(b) = 0. \end{aligned}$$

and hence  $\frac{P(a+h) - P(a)}{h} = 0$ . [See Figure \*\*\*.]

Now for any  $h$  where  $k \neq 0$ , see Figure \*\*\*, the original argument of the pseudoproof is still valid to show that  $\frac{P(a+h) - P(a)}{h}$  approaches 0 as  $h \rightarrow 0$ .

[This is primarily because  $g'(a) = 0$ .]

In summary then, as  $h$  approaches 0 either  $\frac{P(a+h) - P(a)}{h}$  is close to or actually is 0.

$$\text{Thus } P'(a) = \lim_{h \rightarrow 0} \frac{P(a+h) - P(a)}{h} = 0. \quad \text{EOP.}$$

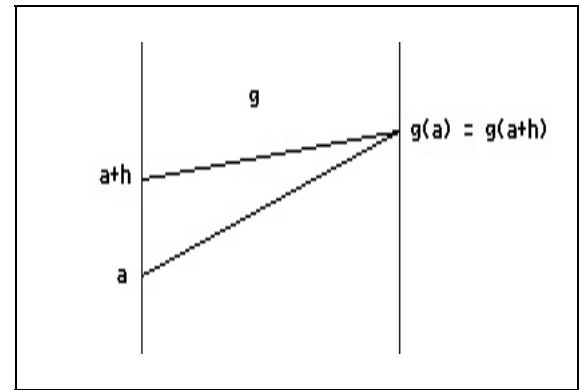


Figure 4

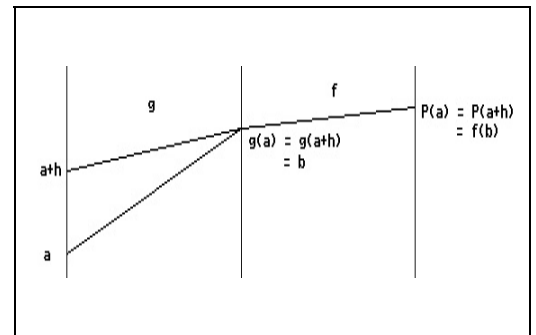


Figure 5