Chapter 9: Partial Derivatives

Learning Objectives:

1. *Describe functions of several variables.*
2. *Define partial derivatives.*
3. *Use Python to evaluate partial derivatives.*
4. *Apply partial derivatives in gradient descent to find the minimum of a function.*

9.1 Partial Derivatives

In Chapter 6, you have been introduced to the notion of a derivative which measures the rate of change of $f(x)$ with respect to the independent variable x. In this chapter, we would extend it to finding the derivatives of function of two or more variables.

The following are examples of functions of two or more variables.

Volume of a cylinder, $V = \pi r^2 h$

Ideal gas law, $PV = \mathrm{n}RT$

Ohms law, $V = IR$

Cobb-Douglas production function, $Q = AL^{\alpha}K^{\beta}$

Let $f(x, y)$ be a function of the two variables x and y. To find the rate of change of $f(x, y)$ w.r.t. both x and y, the technique called **partial differentiation** will be involved. Since $f(x, y)$ is dependent on two variables, we have to, first of all, determine how $f(x, y)$ changes with x while keeping y constant, and how $f(x, y)$ changes with y while keeping x constant. Summing up the effects, the rate of change of $f(x, y)$ w.r.t. both x and y can be evaluated.

Notation for partial derivatives:

The *partial derivative of* $f(x, y)$ *with respect to* x is written as $\dfrac{\partial f}{\partial x}$.

$\dfrac{\partial f}{\partial x}$ is the derivative of $f(x, y)$, where y is treated as the constant and $f(x, y)$ is treated as a function of x alone.

The *partial derivative of* $f(x, y)$ *with respect to* y is written as $\dfrac{\partial f}{\partial y}$.

$\dfrac{\partial f}{\partial y}$ is the derivative of $f(x, y)$, where x is treated as the constant and $f(x, y)$ is treated as a function of y alone.

Example 1.

Compute $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ of the following.

(a) $f(x,y)=5x^3y^2$

(b) $f(x,y)=(4x+3y-5)^8$

(c) $f(x,y)=\sin\left(\frac{x}{1+y}\right)$

[Ans: (a) $\frac{\partial f}{\partial x}=15x^2y^2$, $\frac{\partial f}{\partial y}=10x^3y$ (b) $\frac{\partial f}{\partial x}=32(4x+3y-5)^7$, $\frac{\partial f}{\partial y}=24(4x+3y-5)^7$

(c) $\frac{\partial f}{\partial x}=\frac{1}{1+y}\cos\left(\frac{x}{1+y}\right)$, $\frac{\partial f}{\partial y}=-\frac{x}{(1+y)^2}\cos\left(\frac{x}{1+y}\right)$]

Example 2.

Given that $f(a,b)=\dfrac{a-b}{a+b}$. Evaluate $\dfrac{\partial f}{\partial a}$ and $\dfrac{\partial f}{\partial b}$ when $a = 1$ and $b = 1$.

[Ans: $\left.\dfrac{\partial f}{\partial a}\right|_{a=1,b=1}=\dfrac{1}{2}$, $\left.\dfrac{\partial f}{\partial b}\right|_{a=1,b=1}=-\dfrac{1}{2}$]

9.2 Using SymPy for Partial Derivatives

In Chapter 6, we have learnt to use the SymPy library to obtain the derivative, $\dfrac{dy}{dx}$, of a function, $y=f(x)$. Partial derivatives can also be obtained using the SymPy library. Similar as before, the symbols of the variables need to be declared first.

```python
import sympy as sp

# Define the symbols
x = sp.Symbol('x')
y = sp.Symbol('y')
```

Figure 9.1 Python code for declaring symbols for variables

We then use ***diff(func,var)***, where ***func*** is the expression of the function to be differentiated and ***var*** is the variable we are differentiating with respect to. Since there are more than one variable in the function, thus it is important to indicate ***var*** correctly.

The following code gives the partial derivatives $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ for $z=x^2y+e^{3xy}$:

```python
partial_x = sp.diff(x**2*y+sp.exp(3*x*y),x)
partial_y = sp.diff(x**2*y+sp.exp(3*x*y),y)
print("with respect to x: "+str(partial_x))
print("with respect to y: "+str(partial_y))
```

```
with respect to x: 2*x*y + 3*y*exp(3*x*y)
with respect to y: x**2 + 3*x*exp(3*x*y)
```

Figure 9.2 Python code and output for partial differentiation

From the output, $\dfrac{\partial z}{\partial x}=2xy+3ye^{3xy}$ and $\dfrac{\partial z}{\partial y}=x^2+3xe^{3xy}$.

Example 3.

Find the partial derivatives of the following functions using SymPy.

(a) $f(x,y)=x^2y^3+2\sin x+4y$ (b) $f(x,y)=xe^{xy}$

(c) $f(x,y)=(2x+3y)(5x-4y)$ (d) $f(x,y)=\dfrac{e^{x^2}}{4xy^2}$

[Ans: (a) $f_x(x,y)=2xy^3+2\cos x$, $f_y(x,y)=3x^2y^2+4$ (b) $f_x(x,y)=xye^{xy}+e^{xy}$, $f_y(x,y)=x^2e^{xy}$

(c) $f_x(x,y)=20x+7y$, $f_y(x,y)=7x-24y$ (d) $f_x(x,y)=\dfrac{e^{x^2}}{2y^2}-\dfrac{e^{x^2}}{4x^2y^2}$, $f_y(x,y)=-\dfrac{e^{x^2}}{2xy^3}$]

9.3 An Application of Partial Derivatives – Gradient Descent

Recall in Chapter 8 that you have learnt univariate gradient descent to find the local minimum of a function, $y=f(x)$, of a single variable. In this section, we are going to extend the method to find the local minimum of a function, $z=f(x,y)$, of two variables.

To find the minimum point of a function $z=f(x,y)$, the gradient descent algorithm works in this way:

Algorithm

Step 0. Set a learning rate $\alpha>0$ and an initial point $x=x_0,\ y=y_0$, and compute $f(x_0,y_0)$.

Step 1. At n-th point $x=x_n,\ y=y_n$, compute $f_x(x_n,y_n)$ and $f_y(x_n,y_n)$.

Step 2. Update to the (n+1)-th point using $x_{n+1}=x_n-\alpha f_x(x_n,y_n)$, $y_{n+1}=y_n-\alpha f_y(x_n,y_n)$ and compute $f(x_{n+1},y_{n+1})$.

Step 3. Repeat steps 1 and 2 until a stopping criterion is reached.

Each or a combination of the following can be the stopping criterion:

1. The **maximum number of iterations** is reached.
2. The **value of** $f_x^2(x_n,y_n)+f_y^2(x_n,y_n)$ is smaller than a fixed constant.
3. **Convergence**, which, in simple terms, means that the update to the current point does not differ much from the previous point. It can refer to little difference from $x=x_n,\ y=y_n$ to $x=x_{n+1},\ y=y_{n+1}$ or little reduction in $f(x_n,y_n)$ to $f(x_{n+1},y_{n+1})$ (the difference is smaller than a fixed constant).

Example 4.

To find the minimum point of $f(x,y)=x^2+3y^2$, show the first 3 iterations of the gradient descent algorithm, with an initial point of $x=4,\ y=5$ and step size $\alpha=0.1$.

Solution

We first work out the partial derivatives of $f(x,y)$:

$f_x(x,y)=$ $\qquad$ $f_y(x,y)=$

Iteration 1: Initial point $x_0=4,\ y_0=5 \ \Rightarrow\ f(x_0,y_0)=f(4,5)=$

$f_x(x_0,y_0) = f_x(4,5)$
$=$

$f_y(x_0,y_0) = f_y(4,5)$
$=$

$x_1 = x_0-\alpha f_x(x_0,y_0)$
$=$

$y_1 = y_0-\alpha f_y(x_0,y_0)$
$=$

$f(x_1,y_1)=$

Iteration 2: $x_1=$ $\qquad$, $y_1=$

$f_x(x_1,y_1) =$
$=$

$f_y(x_1,y_1) =$
$=$

$x_2 = x_1-\alpha f_x(x_1,y_1)$
$=$

$y_2 = y_1-\alpha f_y(x_1,y_1)$
$=$

$f(x_2,y_2)=$

Iteration 3: $x_2=$ $\qquad$, $y_2=$ $\qquad$,

$f_x(x_2,y_2) =$
$=$

$f_y(x_2,y_2) =$
$=$

$x_3 = x_2-\alpha f_x(x_2,y_2)$
$=$

$y_3 = y_2-\alpha f_y(x_2,y_2)$
$=$

$f(x_3,y_3)=$

The minimum point of the function $f(x,y)=x^2+3y^2$ is at $x=0,\ y=0$. You can observe that with each iteration, we become nearer to the minimum point.

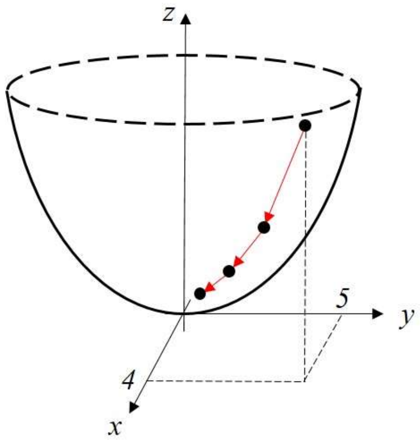

Figure 9.3

[Ans: Iteration 1: $x_1=3.2,\ y_1=2,\ f(x_1,y_1)=22.24$

Iteration 2: $x_2=2.56,\ y_2=0.8,\ f(x_2,y_2)=8.4736$

Iteration 3: $x_3=2.048,\ y_3=0.32,\ f(x_3,y_3)=4.5015$]

Using the function in Example 4, gradient descent can be implemented in Python as follows:

```python
import numpy as np

next_x = 4 # Initial point
next_y = 5 # Initial point
alpha = 0.1 # Learning rate
epsilon = 0.001 # Stopping criterion constant
max_iters = 500 # Maximum number of iterations

# Partial derivatives and function
partialf_x = lambda x,y: 2*x
partialf_y = lambda x,y: 6*y
func = lambda x,y: x**2+3*y**2

next_func = func(next_x,next_y) # Initial value of function

for n in range(max_iters):
    current_x = next_x
    current_y = next_y
    current_func = next_func
    next_x = current_x-alpha*partialf_x(current_x,current_y) # update of x
    next_y = current_y-alpha*partialf_y(current_x,current_y) # update of y
    next_func = func(next_x,next_y)
    change_func = abs(next_func-current_func) # stopping criterion: values of function converge
    print("Iteration",n+1,": x = ",next_x,", y = ",next_y,", f(x,y) = ",next_func)
    if change_func<epsilon:
        break
```

Figure 9.4 Gradient descent code

The convergence criterion used in this code is the difference in the **values of the function**. The output of the code is as follows:

```
Iteration 1 : x =  3.2 , y =  2.0 , f(x,y) =  22.240000000000002
Iteration 2 : x =  2.56 , y =  0.7999999999999998 , f(x,y) =  8.4736
Iteration 3 : x =  2.048 , y =  0.3199999999999999 , f(x,y) =  4.501504
Iteration 4 : x =  1.6384 , y =  0.12799999999999995 , f(x,y) =  2.73350656
Iteration 5 : x =  1.31072 , y =  0.05119999999999997 , f(x,y) =  1.7258512384000002
Iteration 6 : x =  1.0485760000000002 , y =  0.020479999999999984 , f(x,y) =  1.1007699189760005
Iteration 7 : x =  0.8388608000000002 , y =  0.008191999999999993 , f(x,y) =  0.7038887683686403
Iteration 8 : x =  0.6710886400000001 , y =  0.0032767999999999964 , f(x,y) =  0.45039217499176976
Iteration 9 : x =  0.5368709120000001 , y =  0.0013107199999999983 , f(x,y) =  0.28823553011246705
Iteration 10 : x =  0.4294967296000001 , y =  0.0005242879999999992 , f(x,y) =  0.18446826537081645
Iteration 11 : x =  0.3435973836800001 , y =  0.0002097151999999996 , f(x,y) =  0.11805929401313653
Iteration 12 : x =  0.27487790694400005 , y =  8.388607999999985e-05 , f(x,y) =  0.0755578848365376
Iteration 13 : x =  0.21990232555520003 , y =  3.355443199999993e-05 , f(x,y) =  0.0483570361622849
Iteration 14 : x =  0.17592186044416003 , y =  1.3421772799999973e-05 , f(x,y) =  0.030948501522566473
Iteration 15 : x =  0.140737488355328 , y =  5.368709119999989e-06 , f(x,y) =  0.0198070407150352
Iteration 16 : x =  0.11258999068426241 , y =  2.147483647999996e-06 , f(x,y) =  0.012676506016117355
Iteration 17 : x =  0.09007199254740993 , y =  8.589934591999983e-07 , f(x,y) =  0.008112963843674279
Iteration 18 : x =  0.07205759403792794 , y =  3.4359738367999934e-07 , f(x,y) =  0.005192296858889006
Iteration 19 : x =  0.057646075230342354 , y =  1.374389534719997e-07 , f(x,y) =  0.0033230699895189586
Iteration 20 : x =  0.04611686018427388 , y =  5.4975581388799866e-08 , f(x,y) =  0.0021267647932649326
Iteration 21 : x =  0.03689348814741911 , y =  2.1990232555519943e-08 , f(x,y) =  0.0013611294676852048
```

Figure 9.5 Output of the gradient descent code

It can be seen that from Iteration 20 to Iteration 21, the reduction in the value of $f(x, y)$ is: 0.00213 − 0.00136 = 0.00077, which is less than epsilon. Hence, the algorithm converges and the output shows that the minimum point is at $x = 0, y = 0$ and $f(x, y) = 0$.

Example 5.

Repeat the example above to find the minimum of $f(x, y) = x^2 + 3y^2$ with learning rates $\alpha = 0.05$ and $\alpha = 0.5$. Use the difference in values of $f(x, y)$ less than epsilon as the convergence criterion. Set the initial point to be at $x_0 = 4, y_0 = 5$, epsilon = 0.001 and maximum number of iterations to be 500. Compare all results with the above (Example 4) where $\alpha = 0.1$.

Example 6.

Find the minimum point of the function $f(x, y) = x^2 + y^4 + 3y^3 - y^2 - 3y$ using the gradient descent algorithm with the following settings:

(a) Initial point = (1, 1), learning rate = 0.05, epsilon = 0.0001, maximum iteration = 500

(b) Initial point = (0, −3), learning rate = 0.05, epsilon = 0.0001, maximum iteration = 500

Compare and comment on your results.

Tutorial 9

1. Evaluate $f(x,y,z) = x^2 + 2y + z$ when $x = 3,\ y = 4$ and $z = 1$.

2. The volume V of a right circular cylinder is given by the formula $V = \pi r^2 h$, where r is the radius and h is the height.
   (a) Find the formula for the instantaneous rate of change of V with respect to r if r changes and h remains constant.
   (b) Find the formula for the instantaneous rate of change of V with respect to h if h changes and r remains constant.
   (c) Suppose that h has a constant value of 4 cm. but r varies. Find the rate of change of V with respect to r at the point where $r = 6$ cm.
   (d) Suppose that r has a constant value of 8 cm but h varies. Find the instantaneous rate of change of V with respect to h at the point where $h = 10$cm.

3. Find the partial derivatives of the given function.
   (a) $f(x,y) = x^5 + x^3y^2 + 3xy^4$
   (b) $f(r,s) = r \cdot \ln\left(r^2 + s^2\right)$
   (c) $u = te^{w/t}$
   (d) $z = \left(1 + x^2 y\right)e^{3y}$
   (e) $z = \ln(xy)$
   (f) $f(x,y) = x^2 \sin(xy) - 3y^3$
   (g) $g(x,y,z) = xyz + y^2z^3 - 4xz^2$
   (h) $f(x,y,z) = x\cos yz + 2x^2e^z$

4. Evaluate the indicated partial derivatives.
   (a) $f(x,y) = \sqrt{x^2 + y^2}$, $f_x(3,4)$
   (b) $f(x,y,z) = \dfrac{x}{y+z}$, $f_z(3,2,1)$

5. To find the minimum point of $f(x,y) = e^{2x^2+(y-1)^2}$, show the first 2 iterations of the gradient descent algorithm with an initial point of $x = 1,\ y = 2$ and learning rate $\alpha = 0.01$. Using epsilon = 0.001, find the minimum point of $f(x,y)$.
(Use the difference in values of $f(x,y)$ less than epsilon as the convergence criterion.)

6. To find a minimum point of $f(x,y) = \cos(x^2) + y^2$, show the first 3 iterations of the gradient descent algorithm with an initial point of $x = -1,\ y = 2$ and learning rate $\alpha = 0.1$. Find the minimum point of $f(x,y)$. Using epsilon = 0.0001, find the minimum point of $f(x,y)$.
(Use the difference in values of $f(x,y)$ less than epsilon as the convergence criterion.)

7. To find the minimum point of $f(x, y) = (6 - x - 3y)^2 + (2 - x - y)^2$, show the first 3 iterations of the gradient descent algorithm with an initial point of $x = 0,\ y = 0$ and learning rate $\alpha = 0.05$. Using epsilon = 0.00001, find the minimum point of $f(x, y)$. (Use the difference in values of $f(x, y)$ less than epsilon as the convergence criterion.)

8.* Use the gradient descent algorithm to find the **maximum** point of $f(x, y) = e^{-x^2+2x} - y^2$. Set the initial point at $x = 2,\ y = 1$, learning rate $\alpha = 0.05$ and epsilon = 0.00001. Use the difference in values of $f(x, y)$ less than epsilon as the convergence criterion.
(Hint: Find the minimum point of $g(x, y) = -f(x, y)$.)

9. Use the gradient descent algorithm to find the minimum point of $f(x, y) = \cos 0.125x^3 - \sin y$ with the following settings:
   (a) Initial point: $x = 2.5,\ y = 1$, learning rate $\alpha = 0.1$
   (b) Initial point: $x = -2,\ y = -4$, learning rate $\alpha = 0.1$
   (c) Initial point: $x = -2,\ y = -4$, learning rate $\alpha = 0.01$
   (d) Initial point: $x = -2,\ y = -4$, learning rate $\alpha = 0.5$

   Set epsilon = 0.0001 for all three cases and 500 as maximum number of iterations. Use the difference in values of $f(x, y)$ less than epsilon as the convergence criterion. Compare and comment on your results.

Answers

1. 18

2. (a) $2\pi rh$ (b) πr^2 (c) 48π (d) 64π

3. (a) $f_x(x,y) = 5x^4 + 3x^2y^2 + 3y^4$; $f_y(x,y) = 2x^3y + 12xy^3$

   (b) $f_r(r,s) = \ln\left(r^2+s^2\right) + \dfrac{2r^2}{r^2+s^2}$; $f_s(r,s) = \dfrac{2rs}{r^2+s^2}$

   (c) $\dfrac{\partial u}{\partial t} = e^{\frac{w}{t}}\left(1-\dfrac{w}{t}\right)$; $\dfrac{\partial u}{\partial w} = e^{\frac{w}{t}}$

   (d) $\dfrac{\partial z}{\partial x} = 2xye^{3y}$; $\dfrac{\partial z}{\partial y} = x^2e^{3y} + 3e^{3y}\left(1+x^2y\right)$

   (e) $\dfrac{\partial z}{\partial x} = \dfrac{1}{x}$; $\dfrac{\partial z}{\partial y} = \dfrac{1}{y}$

   (f) $f_x(x,y) = 2x\sin(xy) + x^2y\cos(xy)$; $f_y(x,y) = x^3\cos(xy) - 9y^2$

   (g) $g_x(x,y,z) = yz - 4z^2$; $g_y(x,y,z) = xz + 2yz^3$; $g_z(x,y,z) = xy + 3y^2z^2 - 8xz$

   (h) $f_x(x,y,z) = \cos yz + 4xe^z$; $f_y(x,y,z) = -xz\sin yz$; $f_z(x,y,z) = -xy\sin yz + 2x^2e^z$

4. (a) $\dfrac{3}{5}$ (b) $-\dfrac{1}{3}$

5. Iteration 1: $x_1 = 0.1966, y_1 = 1.5983, f(x_1, y_1) = 1.5453$

   Iteration 2: $x_2 = 0.1844, y_2 = 1.5798, f(x_2, y_2) = 1.4981$

   ... Iteration 61: $x = 0.0119, y = 1.150, f(x,y) = 1.023$

6. Iteration 1: $x_1 = -1.1683, y_1 = 1.6, f(x_1, y_1) = 2.7644$

   Iteration 2: $x_2 = -1.3970, y_2 = 1.28, f(x_2, y_2) = 1.2667$

   Iteration 3: $x_3 = -1.6564, y_3 = 1.024, f(x_3, y_3) = 0.1267$

   … Iteration 23: $x = -1.772, y = 0.0118, f(x,y) = -1.0$

7. Iteration 1: $x_1 = 0.8, y_1 = 2, f(x_1, y_1) = 1.28$

   Iteration 2: $x_2 = 0.64, y_2 = 1.68, f(x_2, y_2) = 0.2048$

   Iteration 3: $x_3 = 0.64, y_3 = 1.744, f(x_3, y_3) = 0.16384$

   … Iteration 105: $x = 0.0180, y = 1.993, f(x,y) = 0.000131$

8*. Iteration 48: $x = 1.0, y = 0.00636, f(x,y) = 2.718$

9. (a) Iteration 29: $x = 2.929, y = 1.543, f(x,y) = -2.0$

   (b) Iteration 31: $x = -2.929, y = -4.684, f(x,y) = -2.0$

   (c) Iteration 201: $x = -2.929, y = -4.643, f(x,y) = -2.0$

   (d) Did not converge after 500 iterations

   Different initial points may lead to different minimum points as evident in (a) and (b). A lower learning rate may result in slower convergence, comparing (b) and (c). Too large a learning rate may result in the algorithm to fail to converge as seen in (d).