Class 8 Notes: Partial Derivatives and Multiple Integrals

Myles D. Garvey, Ph.D

Summer, 2019

Contents

1 Introduction

In the last lecture, we covered the fundamentals of multivariable functions. Now that we have an understanding of how to define and visualize these, we will turn our attention to multivariable calculus. This topic sounds very scary to many people. "Calculus" alone scares many whom have not ventured into the study of limits, derivatives, and integrals. However, if you have managed to master these three topics, multi-variable calculus is actually quite simple. It is merely an extension of our concepts to functions of multiple variables. Many of our equations remain the same, with only small modifications. However, as we will see, multivariable calculus is when we combine some of the elements of Linear Algebra (ie Matrix Algebra) and ordinary calculus. In this lecture, we will review through two important topics, namely partial derivatives and multiple integrals.

2 Partial Derivatives

2.1 Definition of a Partial Derivative

In the last lecture, we defined a limit of a multivariable function as the result of a function that approaches a number L as a result of taking a point x in n-dimensional space "really close" to another point a in n-dimensional space. It would hence be natural to extend our definition of derivatives to functions of multiple variables, since we now have a way to compute limits of multivariable functions.

Unlike ordinary calculus, however, where we can bring only a single dimension of space (x) "close to" a point a, we now have the problem of bringing multiple dimensions of space "close to" a point a. In single variable calculus, we defined derivatives as the slope of a tangent line to the curve $f(x)$ at a point x, as we took the difference between one point and another on the x-axis. In multi-dimensional space, we can bring one point (x_1, y_1) "close to" another point (x_2, y_2) and see how the resulting slope of $f(x, y)$ changes. In 3-dimensional space (that is, the x, y, and $f(x, y)$), we no longer have lines that are tangent to curves. Instead, we have *planes* that are tangent to *surfaces*. We saw the concept of surface in the last lecture when we plotted some of our functions. Therefore, to understand how the values of $f(x, y)$ are changing at a specific point, we not only need to characterize this along a single x dimension, but also the y dimension, since we always need two dimensions to define a plane in $3 - D$ space.

Definition 1 *A plane in 3D space is characterized by the following equation:*
$f(x, y) = a_0 + a_1 x + a_2 y$.
In $n + 1$ dimensional space, a hyper-plane is characterized by the following equation:
$f(x_1, x_2, \ldots, x_n) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$

In general, the a_i is called the slope, or rate of change, of the function along the x_i dimension.

Definition 2 *In order to characterize the rate of change of a multivariable function $f(x_1, \ldots, x_n)$ in $n + 1$ dimensional space at a point $(x_1, \ldots, x_n)$, we need to understand the slopes along each dimension of the plane that is tangent to the function at the point $(x_1, \ldots, x_n)$.*

In order to understand rate of change of the function, we need to select a dimension to reference the rate of change to. For example, we can have the rate of change of a cost function with respect to the change in inventory holding cost. But we can also have the rate of change of the same cost function with respect to the change in fixed ordering cost. That is, there is not more concept of only a single "slope", but rather, each slope is *with respect to* a given dimension x_i. If we were to have a function $f(x_1, x_2, \ldots, x_n)$, suppose we set a value for each $x_2 = a_2, x_3 = a_3, \ldots, x_n = a_n$. Then we effectively collapse the function into a single variable function: $f(x_1) = f(x_1, a_2, \ldots, a_n)$. That is, if we assign numbers to each of the other dimensions, then the multivariable function is converted into a single variable function, and we can then apply calculus on the *cross-section* of the function. Rather than assigning numbers to each of the other dimensions, we just simply assume that the other variables are numbers. When we do this, we then can use ordinary claculus to define a specfic type of derivative:

Definition 3 *Let $f(x_1, \ldots, x_n)$ by a function of n variables. Pick a dimension i and a point $(a_1, a_2, \ldots, a_n)$. Assume that all variables x_j, with $j \neq i$ are set equal to a number. Then the derivative of the resulting single variable function is called the partial derivative of f with respect to x_i at point $(a_1, \ldots, a_n)$, and is defined as:*

$$f_{x_i}(a_1, \ldots, a_n) = \frac{\partial f}{\partial x_i} = D_{x_i} f = \lim_{\Delta x_i \to 0} \frac{f(a_1, \ldots, a_i + \Delta x_i, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{\Delta x_i}$$

Example 1 *Consider $f(x_1, x_2) = x_1 x_2$. We can find the partial derivative with respect to x_1 at point (a_1, a_2) by assuming that $x_2 = a_2$. Doing so gives us $f(x_1, a_2) = x_1 a_2 = a_2 x_1$. Using the equation for the partial derivative, we have*

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \lim_{\Delta x_1 \to 0} \frac{f(a_1 + \Delta x_1, a_2) - f(a_1, a_2)}{\Delta x_1} \\
&= \lim_{\Delta x_1 \to 0} \frac{(a_1 + \Delta x_1)(a_2) - (a_1)(a_2)}{\Delta x_1} \\
&= \lim_{\Delta x_1 \to 0} \frac{a_1 a_2 + \Delta x_1 a_2 - a_1 a_2}{\Delta x_1} \\
&= \lim_{\Delta x_1 \to 0} \frac{\Delta x_1 a_2}{\Delta x_1} \\
&= \lim_{\Delta x_1 \to 0} a_2 \\
&= a_2
\end{aligned}$$

Likewise, we can find the partial derivative with respect to x_2:

$$\begin{aligned}
\frac{\partial f}{\partial x_2} &= \lim_{\Delta x_2 \to 0} \frac{f(a_1, a_2 + \Delta x_2) - f(a_1, a_2)}{\Delta x_2} \\
&= \lim_{\Delta x_2 \to 0} \frac{(a_1)(a_2 + \Delta x_2) - (a_1)(a_2)}{\Delta x_2} \\
&= \lim_{\Delta x_2 \to 0} \frac{a_1 a_2 + \Delta x_2 a_1 - a_1 a_2}{\Delta x_2} \\
&= \lim_{\Delta x_2 \to 0} \frac{\Delta x_2 a_1}{\Delta x_2} \\
&= \lim_{\Delta x_1 \to 0} a_1 \\
&= a_1
\end{aligned}$$

2.2 Rules of Partial Differentiation

Working with and finding partial derivatives works very simliar, if not the same, as working with ordinary derivatives. Partial derivatives can be easily found using the same rules that we derived in the previous lecture. Recall that when we take a partial derivative, we are essentially assuming that all other variables are constants (this was clearly illustrated in the example above). If this is the case, then finding a partial derivative with respect to a given variable is just simply a matter of doing regular old differentiation, but simply assuming that all other variables in the equation are held constant.

Theorem 1 *Let $f(x_1, \ldots, x_n)$ be a function of n variables. Then the partial derivative of f can be found if f is one of the following types of functions:*

$$\frac{\partial}{\partial x_i}[af(x_1,\ldots,x_n)] = a\frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)], a \in \mathbb{R} \tag{1}$$

$$\frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)+g(x_1,\ldots,x_n)] = \frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)] + \frac{\partial}{\partial x_i}[g(x_1,\ldots,x_n)] \tag{2}$$

$$\frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)-g(x_1,\ldots,x_n)] = \frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)] - \frac{\partial}{\partial x_i}[g(x_1,\ldots,x_n)] \tag{3}$$

$$\frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)g(x_1,\ldots,x_n)] = g(x_1,\ldots,x_n)\frac{\partial}{\partial x_i}[f(x_1,\ldots,x_n)] + f(x_1,\ldots,x_n)\frac{\partial}{\partial x_i}[g(x_1,\ldots,x_n)] \tag{4}$$

$$\frac{\partial}{\partial x_i}[a] = 0, a \in \mathbb{R} \tag{5}$$

$$\frac{\partial}{\partial x_i}[x_i^n] = nx^{n-1} \tag{6}$$

$$\frac{\partial}{\partial x_i}[a^{x_i}] = a^{x_i}ln(a) \tag{7}$$

$$\frac{\partial}{\partial x_i}[\ln(x_i)] = \frac{1}{x_i} \tag{8}$$

$$\frac{\partial}{\partial x_i}[\log_a(x_i)] = \frac{1}{x_i\ln(a)} \tag{9}$$

$$\frac{\partial}{\partial x_i}[e^{x_i}] = e^{x_i} \tag{10}$$

Example 2 *Let $f(x_1, x_2) = x_1^2 + x_1e^{-x_1x_2}$. Find $\frac{\partial f}{\partial x_1}and\frac{\partial f}{\partial x_2}$. We can see that for $\frac{\partial f}{\partial x_1}f(x_1, x_2)$, we hold x_2 constant and treat this as a function of one variable (namely a function of x_1). Doing so leads us to:*

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= \frac{\partial}{\partial x_1}[x_1^2 + x_1 e^{-x_1 x_2}] \\
&= \frac{\partial}{\partial x_1}[x_1^2] + \frac{\partial}{\partial x_1}[x_1 e^{-x_1 x_2}] \\
&= 2x_1 + \frac{\partial}{\partial x_1}[x_1]e^{-x_1 x_2} + x_1 \frac{\partial}{\partial x_1}[e^{-x_1 x_2}] \\
&= 2x_1 + e^{-x_1 x_2} + x_1 e^{-x_1 x_2} \frac{\partial}{\partial x_1}[-x_1 x_2] \\
&= 2x_1 + e^{-x_1 x_2} - x_1 e^{-x_1 x_2} x_2
\end{aligned}
$$

On the other hand, if we consider x_1 to be a constant and we differentiate with respect to x_2 instead, we obtain:

$$
\begin{aligned}
\frac{\partial f}{\partial x_2} &= \frac{\partial}{\partial x_2}[x_1^2 + x_1 e^{-x_1 x_2}] \\
&= \frac{\partial}{\partial x_2}[x_1^2] + \frac{\partial}{\partial x_2}[x_1 e^{-x_1 x_2}] \\
&= (0) + x_1 \frac{\partial}{\partial x_2}[e^{-x_1 x_2}] \\
&= x_1 e^{-x_1 x_2} \frac{\partial}{\partial x_2}[(-x_1 x_2)] \\
&= -x_1^2 e^{-x_1 x_2}
\end{aligned}
$$

2.3 Higher-Order Partial Derivatives

Just like in single variable calculus, we can repeatedly take derivatives of derivatives. However, unlike single variable calculus, we can take derivatives of functions with respect to one variable, and then subsequently take the derivative of the resulting variable with respect to a different variable (or the same one). To illustrate the process, we use the following example:

Example 3

$$
\begin{aligned}
\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_2}[x_1 x_2^2 + x_2 x_1^2] &= \frac{\partial f}{\partial x_1}\left[\frac{\partial f}{\partial x_2}[x_1 x_2^2 + x_2 x_1^2]\right] \\
&= \frac{\partial f}{\partial x_1}\left[[2x_1 x_2 + x_1^2]\right] \\
&= 2x_2 + 2x_1
\end{aligned}
$$

$$\begin{aligned}\frac{\partial f}{\partial x_2}\frac{\partial f}{\partial x_1}[x_1x_2^2 + x_2x_1^2] &= \frac{\partial f}{\partial x_2}\left[\frac{\partial f}{\partial x_1}[x_1x_2^2 + x_2x_1^2]\right] \\ &= \frac{\partial f}{\partial x_2}\left[x_2^2 + 2x_1x_2\right] \\ &= 2x_2 + 2x_1\end{aligned}$$

$$\begin{aligned}\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_1}[x_1x_2^2 + x_2x_1^2] &= \frac{\partial f}{\partial x_1}\left[\frac{\partial f}{\partial x_1}[x_1x_2^2 + x_2x_1^2]\right] \\ &= \frac{\partial f}{\partial x_1}\left[x_2^2 + 2x_2x_1\right] \\ &= 2x_2\end{aligned}$$

$$\begin{aligned}\frac{\partial f}{\partial x_2}\frac{\partial f}{\partial x_2}[x_1x_2^2 + x_2x_1^2] &= \frac{\partial f}{\partial x_2}\left[\frac{\partial f}{\partial x_2}[x_1x_2^2 + x_2x_1^2]\right] \\ &= \frac{\partial f}{\partial x_2}\left[2x_1x_2 + x_1^2\right] \\ &= 2x_1\end{aligned}$$

Typically, we denote these higher order derivatives as follows:

Definition 4 *Let $f(x_1, \ldots, x_n)$ be a multvariable function of n variables). Then the second derivatives of f are denoted as follows, for all $i, j \in \{1, 2, \ldots, n\}$:*

$$\begin{aligned}f_{x_ix_i} &= \frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left[\frac{\partial f}{\partial x_i}\right] \\ f_{x_ix_j} &= \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}\left[\frac{\partial f}{\partial x_i}\right]\end{aligned}$$

Caution must be made in reading these. Notice how notationally we write $f_{x_ix_j}$. Reading this from left to right, this says to first to the partial derivative with respect to x_i, then take the partial derivative with respect to x_j. Notice that this is more clear when we write $\frac{\partial^2 f}{\partial x_j \partial x_i}$, since the variable all the way to the right would be the first to which we apply the partial derivative. We notice in the example above something interesting, namely that $f_{x_1x_2} = f_{x_2x_1}$. This is no coincidence:

Theorem 2 *Let $f(x_1, x_2)$ be a function of two variables which is continuously differentiable on the open ball B. Then $f_{x_1x_2} = f_{x_2x_1}$. In other words, if we have a function of two variables that is continuously differentiable on the ball B, then the order in which we take second partial derivatives with different variables does not matter.*

3 Fusing Calculus and Linear Algebra

3.1 The Gradient, Jacobian, and Hessian Matrices

At this point, we have covered enough material to slightly move our lecture in a different direction (no pun intended). We have covered the fundamentals of vectors, vector spaces, and matrices. As it turns out, when we are working with derivatives in multi-dimensional space, we can represent the derivatives of functions at points in terms of a vector.

Definition 5 *Let f be a function of n independent variables. Then the gradient of f, denoted as ∇f, is the vector that is formed by finding the partial derivatives of f and multiplying them by standard basis vectors in n dimensional space. In other words, if we have the standard basis vectors*

$\boldsymbol{e_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{e_2} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \ldots, \boldsymbol{e_n} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$*, then the gradient of f is defined as the vector:*

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \frac{\partial f}{\partial x_1}\boldsymbol{e_1} + \frac{\partial f}{\partial x_2}\boldsymbol{e_2} \cdots + \frac{\partial f}{\partial x_n}\boldsymbol{e_n}$$

When the function is a vector valued function, that is, it is a function $f : \mathbb{R} \to \mathbb{R}$, then the matrix comprising of the transpose of each gradient vector of each function is known as the *Jacobian* matrix. We will omit further details from here on this. For now, let us look at an example of finding the gradient of a function:

Example 4 *Let $f(x_1, x_2) = 2x_1^{x_2} - x_1x_2$. Finding the partials of each, we have:*

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= \frac{\partial}{\partial x_1}[2x_1^{x_2} - x_1x_2] \\ &= 2x_2x_1^{x_2-1} - x_2 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial f}{\partial x_2} &= \frac{\partial}{\partial x_2}[2x_1^{x_2} - x_1x_2] \\ &= 2x_1^{x_2}\ln(x_1) - x_1 \end{aligned}$$

Therefore, the gradient vector would be:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_2x_1^{x_2-1} - x_2 \\ 2x_1^{x_2}\ln(x_1) - x_1 \end{bmatrix}$$

In addition to the matrix representations of these derivatives, we can also form another matrix, which is of extreme importance to us in later applications, called the *Hessian Matrix*:

Definition 6 *The Hessian Matrix of a function $f(x_1, \dots, x_n)$ is the matrix of second derivatives, where the rows are the second derivatives with that variable as the first derivative, and the columns representing that variable as the second derivative. That is:*

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

We also know generally that if we have a function of multiple variables that is continuously differentiable in the domain, then

Theorem 3 *Let $f(x_1, \dots, x_n)$ be a continuously differentiable function on it's domain. Then for all $i, j \in \{1, \dots, n\}$, we have:*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Since the elements of the Hessian matrix are $\frac{\partial^2 f}{\partial x_i \partial x_j}$, this would imply that by definition, it is a symmetric matrix. This means we can easily determine if the matrix is *positive definite* or *negative definite*, which helps us with further applications later:

Theorem 4 *Let A be an $n \times n$ matrix. Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}$. Then $\boldsymbol{x}^T A \boldsymbol{x}$ is a number. This should be easy to see, since $\boldsymbol{x}^T$ is an $1 \times n$ matrix, A is an $n \times n$ matrix, and $\boldsymbol{x}$ is a $1 \times n$ matrix. We can see that the first multiplication yields a $(1 \times n)(n \times n) = 1 \times n$ matrix. The multiplication of this with the vector yields an $(1 \times n)(n \times 1) = (1 \times 1)$ matrix. A 1×1 matrix is always interpreted as a number.*

Definition 7 *An $n \times n$ matrix A is said to be positive definite if for all vectors $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}$, $\boldsymbol{x}^T A \boldsymbol{x} > 0$. Likewise, it is said to be negative definite if for all vectors $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}$, $\boldsymbol{x}^T A \boldsymbol{x} < 0$.*

Proving that a matrix is positive definite is sometimes very difficult to accomplish, since it needs to be shown that $\boldsymbol{x}^T A \boldsymbol{x} > 0$ for **all** vectors $\boldsymbol{x}$, not just a handful of them. However, if our matrix is symmetric, there is an easy approach to determining if it is positive or negative definite:

Theorem 5 *A symmetric square $n \times n$ matrix A is positive definite if and only if it's eigenvalues are all positive. Likewise, it is negative definite if and only if it's eigenvalues are all negative.*

3.2 Minimizing and Maximizing Unconstrained Functions

Functions of multiple values also have the concept of maximum and minimum values analogous to single variable calculus. The reason is simple: our codomain of the function is the real number line. Therefore, we also have the concept of minimizing and maximizing functions of several variables:

Definition 8 *Let $f : A_1 \times \cdots \times A_n \to \mathbb{R}$ be a function of n independent variables. We say that M is an absolute maximum at the point $(x_1^*, x_2^*, \ldots, x_n^*)$ if $f(x_1, x_2, \ldots, x_n) \leq M$ for all $(x_1, \ldots, x_n) \in A_1 \times \cdots \times A_n$.*

Definition 9 *Let $f : A_1 \times \cdots \times A_n \to \mathbb{R}$ be a function of n independent variables. We say that m is an absolute minimum at the point $(x_1^*, x_2^*, \ldots, x_n^*)$ if $m \leq f(x_1, x_2, \ldots, x_n)$ for all $(x_1, \ldots, x_n) \in A_1 \times \cdots \times A_n$.*

Given our definitions, we can use an analogous approach to finding the maximum and minimum values of a multivariable function just as we had done in single variable calculus:

Definition 10 *A function is said to be unconstrained if we are maximizing or minimizing the function by taking into account all values in the entire domain of the function. On the other hand, the function is said to be constrained if we are only considering a subset of the domain, of which is often defined by other equations or inequalities that must hold true about the points in n dimensional space in the defined set.*

Definition 11 *Consider an unconstrained function $f(x_1, \ldots, x_n)$ at a point $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)$. Then $\boldsymbol{x}^*$ is said to be a critical or stationary point if $\nabla f = \mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$. In other words, $\boldsymbol{x}^*$ is a critical or stationary point if all of the partial derivatives with respect to each dimension, respectively, is equal to 0.*

Theorem 6 *Consider an unconstrained function $f(x_1, \ldots, x_n)$, $f : A \to \mathbb{R}$ (where A is a Cartesian product of n sets) at a point $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)$, and, consider the Hessian Matrix $H(\boldsymbol{x}^*)$ of this function at this point. If $\nabla f = \mathbf{0}$ and if $H(\boldsymbol{x}^*)$ is positive definite, then $f(x_1^*, \ldots, x_n^*)$ is a minimum, and we write $(x_1^*, \ldots, x_n^*) = \underset{(x_1,\ldots,x_n)\in A}{\arg\min} \{f(x_1, \ldots, x_n)\}$. If $H(\boldsymbol{x}^*)$ is negative definite, then $f(x_1^*, \ldots, x_n^*)$ is a maximum, and we write $(x_1^*, \ldots, x_n^*) = \underset{(x_1,\ldots,x_n)\in A}{\arg\max} \{f(x_1, \ldots, x_n)\}$.*

Example 5 *Consider the function $f(x_1, x_2) = 10 - 2x_1^2 - 2x_2^2$. We would like to find the maximum or minimum if this function has one. First, we find the gradient:*

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}[10 - 2x_1^2 - 2x_2^2] = -2x_1$$
$$\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2}[10 - 2x_1^2 - 2x_2^2] = -2x_2$$

Therefore, we have $\nabla f = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$. To find the critical or stationary point, we solve for $\nabla f = \mathbf{0}$:

$$\nabla f = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix} = \mathbf{0}$$

$$\begin{matrix} -2x_1 & & = 0 \\ & -2x_2 & = 0 \end{matrix} \rightarrow \begin{matrix} x_1 & & = 0 \\ & x_2 & = 0 \end{matrix}$$

Therefore, a potential candidate for a maximum or minimum would be $(x_1^*, x_2^*) = (0,0)$. Now we find the Hessian at the point $(0,0)$:

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1 x_1} &= \frac{\partial}{\partial x_1}\left[\frac{\partial f}{\partial x_1}\right] \\ &= \frac{\partial}{\partial x_1}\left[-2x_1\right] \\ &= -2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1 x_2} &= \frac{\partial}{\partial x_1}\left[\frac{\partial f}{\partial x_2}\right] \\ &= \frac{\partial}{\partial x_1}\left[-2x_2\right] \\ &= 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x_2 x_1} &= \frac{\partial^2 f}{\partial x_1 x_2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x_2 x_2} &= \frac{\partial}{\partial x_2}\left[\frac{\partial f}{\partial x_2}\right] \\ &= \frac{\partial}{\partial x_2}\left[-2x_2\right] \\ &= -2 \end{aligned}$$

Therefore, for any point (x_1, x_2), we have $H((x_1,x_2)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 x_1} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2 x_2} \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$

We use software to find the eigenvalues for this matrix:

```
>>> import numpy as np
>>> x=np.array([[-2,0],[0,-2]])
>>> eval,evec=np.linalg.eig(x)
>>> eval
array([-2.,  -2.])
```

So, we have $\lambda_1 = \lambda_2 = -2$. Since $\nabla f(0,0) = \mathbf{0}$, and $H((0,0))$ is negative definite, we have shown that a maximum exists for the function at $(0,0)$, and we can write $\underset{(x_1,x_2)\in\mathbb{R}\times\mathbb{R}}{\arg\max}\ \{10 - 2x_1^2 - 2x_2^2\} = (0,0)$

Example 6 *Consider the function $f(x_1, x_2) = 3xe^{-5x} + 2ye^{-5y}$. We would like to find the minimum or maximum, if one exists. Recall that we begin by finding $\nabla f = \mathbf{0}$:*

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}[3xe^{-5x} + 2ye^{-5y}] = 3e^{-5x} - 15xe^{-5x}\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}[3xe^{-5x} + 2ye^{-5y}] = 2e^{-5y} - 10ye^{-5y}$$

Hence, $\nabla f = \begin{bmatrix} 3e^{-5x} - 15xe^{-5x} \\ 2e^{-5y} - 10ye^{-5y} \end{bmatrix}$*. Solving for* $\mathbf{0}$*, we have:*

$$\begin{aligned} \nabla f &= \mathbf{0} \\ \begin{bmatrix} 3e^{-5x} - 15xe^{-5x} \\ 2e^{-5y} - 10ye^{-5y} \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \textit{First, solve for } x : 3e^{-5x} - 15xe^{-5x} &= 0 \\ 3e^{-5x} &= 15xe^{-5x} \\ x &= \frac{3}{15} = \frac{1}{5} \\ \textit{Now, solve for } y : 2e^{-5y} - 10ye^{-5y} &= 0 \\ 2e^{-5y} &= 10ye^{-5y} \\ y &= \frac{2}{10} = \frac{1}{5} \end{aligned}$$

Therefore, a critical point is $(\frac{1}{5}, \frac{1}{5}$*. Now we find the Hessian for any point by finding the second partial derivatives:*

$$\begin{aligned} \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right] &= \frac{\partial}{\partial x}\left[3e^{-5x} - 15xe^{-5x}\right] = (-5)3e^{-5x} + 75xe^{-5x} - 15e^{-5x} = 75xe^{-5x} - 30e^{-5x} \\ \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial y}\right] &= \frac{\partial}{\partial y}\left[3e^{-5x} - 15xe^{-5x}\right] = 0 \\ \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial y}\right] &= \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial x}\right] = 0 \\ \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial y}\right] &= \frac{\partial}{\partial y}\left[2e^{-5y} - 10ye^{-5y}\right] = (-5)2e^{-5y} + 50ye^{-5y} - 10e^{-5y} = 50ye^{-5y} - 20e^{-5y} \end{aligned}$$

Now we can form the Hessian for any (x, y):

$$H(x,y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x \partial x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y \partial y} \end{bmatrix} = \begin{bmatrix} 75xe^{-5x} - 30e^{-5x} & 0 \\ 0 & 50ye^{-5y} - 20e^{-5y} \end{bmatrix}$$

Now we can find the Hessian for the specific critical point $(\frac{1}{5}, \frac{1}{5}$*:*

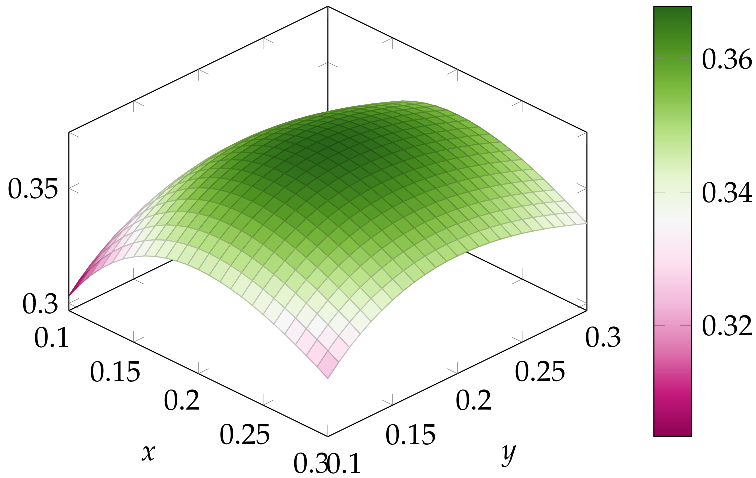

Figure 1: The plot of $3 * x * e^{-5*x} + 2 * y * e^{-5*y}$ near the maximum point at $(\frac{1}{2}, \frac{1}{2})$.

$$
\begin{aligned}
H(\frac{1}{5}, \frac{1}{5}) &= \begin{bmatrix} 75\left(\frac{1}{5}\right)e^{-5\left(\frac{1}{5}\right)} - 30e^{-5\left(\frac{1}{5}\right)} & 0 \\ 0 & 50\left(\frac{1}{5}\right)e^{-5\left(\frac{1}{5}\right)} - 20e^{-5\left(\frac{1}{5}\right)} \end{bmatrix} \\
&= \begin{bmatrix} 15e^{-1} - 30e^{-1} & 0 \\ 0 & 10e^{-1} - 20e^{-1} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{15}{e} & 0 \\ 0 & -\frac{10}{e} \end{bmatrix}
\end{aligned}
$$

Now, putting this matrix in Python and finding the Eigenvalues, we are left with:

```
>>> import numpy as np
>>> import math
>>> x=np.array([[-15.*math.exp(-1,),0.],[0.,-10.*math.exp(-1.)]])
>>> x
array([[-5.51819162,  0.        ],
       [ 0.        , -3.67879441]])
>>> eval,evec=np.linalg.eig(x)
>>> eval
array([-5.51819162, -3.67879441])
```

We notice that both eigenvalues are negative, which means our Hessian (since it is symmetric) is negative definite. Therefore, the function has a point at $(\frac{1}{5}, \frac{1}{5})$*. We can also see this illustrated in Figure 1.*