Today we properly discuss the second derivative test and introduce constrained optimization and the method of Lagrange multipliers.

1. **The Second Derivative Test**

Recall the form of the second derivative test for scalar valued functions of many variables. Remember that the form of the second derivative test given in the book only dealt with functions of two variables. We want something similarly strong for functions of $n$ variables, but unfortunately the technique we used for two variables will not work in general. Still, our starting point is the fundamental

**Theorem 1.1. (Second Derivative Test, primitive form)** Suppose that function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ has continuous second partial derivatives near a critical point $a$ (note that this means $\nabla f(a) = \mathbf{0}$ since by assumption $f$ is differentiable at $a$). Let $u$ be a unit vector based at $a$, and let $g_u(h)$ be the “cross section function” appearing in the definition of the directional derivative

$$g_u(h) = f(a + hu)$$

If

$$g''_u(0) < 0$$

for all directions $u$, then $a$ is a local maximum of $f$; if

$$g''_u(0) > 0$$

for all directions $u$, then $u$ is a local minimum of $f$. If there is a direction $u$ such that

$$g''_u(0) = 0$$

then the test is inconclusive.

First, note that the proof of this theorem is very, very easy. The idea is that by restricting to the “cross section function” we are looking at $f$, but only in the direction of $u$. So if this single variable function $g_u(h)$ has a maximum at $h = 0$ for all directions $u$, then $f$ has a maximum at $a$. More intuitively, if you are on a hill and if in every direction the hill is sloping down, then you are at the top.

Still, while this theorem in theory gives a way to test whether critical points are maxima or minima, there are some problems. In practice, the test is in this form is quite useless, since it requires us to check a condition on $g''_u(0)$ for every $u$, which is infeasible. So we would prefer to have, instead of many different object (all of the $g''_u(0)$) to deal, a single object which encapsulates the behavior of all of the $g''_u$. This single object is the second derivative matrix, or Hessian

$$D^2 f$$

Using the Hessian, we will be able to rewrite the second derivative test above in a usable form.

1.1. **The Hessian.** Recall that given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the gradient $\nabla f$ is a map $\mathbb{R}^n \rightarrow \mathbb{R}^n$, given by

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$
If we differentiate this map, we end up with the matrix of all second partial derivatives

\[
D(\nabla f) = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\
\frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x)
\end{bmatrix}
\]

This is what we call the second derivative or Hessian of \( f \), notated \( D^2 f \).

**Example 1.2.** Find the Hessian of \( f \) at \((0, 0)\) when

\[
f(x, y) = \ln(e^x + e^y)
\]

We compute

\[
fx = \frac{1}{e^x + e^y}e^x
\]

\[
fy = \frac{1}{e^x + e^y}e^y
\]

and that

\[
f_{xx} = e^x e^y \frac{1}{(e^x + e^y)^2}
\]

\[
f_{xy} = -e^x e^y \frac{1}{(e^x + e^y)^2}
\]

\[
f_{yx} = -e^x e^y \frac{1}{(e^x + e^y)^2}
\]

\[
f_{yy} = e^x e^y \frac{1}{(e^x + e^y)^2}
\]

So that

\[
D^2 f = e^x e^y \frac{1}{(e^x + e^y)^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}
\]

so

\[
D^2 f|_{x=0} = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}
\]

In general, when \( f \) has continuous second partial derivatives, the Hessian happens to fall into a class of matrices with very nice properties: the symmetric matrices. So we digress for a moment and discuss symmetric matrices.

### 1.2. Symmetric Matrices

We define the transpose of an \( m \times n \) matrix \( A = [a_{ij}]_{i,j} \) to be the \( n \times m \) matrix \( tA = [a_{ji}]_{i,j} \). That is, \( tA \) is the matrix which switches the \((i, j)\)-th entry of \( A \) with the \((j, i)\)-th entry. So, for example, if \( A \) is the \( 2 \times 3 \) matrix

\[
A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}
\]

then the transpose of \( A \) is a \( 3 \times 2 \) matrix

\[
tA = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}
\]

**Remark 1.3.** The transpose of an \( n \times 1 \) matrix, i.e. a column vector, is a \( 1 \times n \) matrix, i.e. a row (co)vector.

**Remark 1.4.** \( t(tA) = A \) always. Note too that \( t(AB) = tB^tA \), i.e. if \( AB \) makes sense, then \( tB^tA \) makes sense as well, and \( t(AB) = tB^tA \).
Definition 1.5. A matrix $A$ is said to be symmetric if $A^T = A$.

Remark 1.6. Every symmetric matrix $A$ must be square, i.e. must be $n \times n$ (have the same number of rows as columns). This is very easy to see.

The source of our interest in symmetric matrices comes from the following important (but easy) observation.

Theorem 1.7. (Clairaut’s Theorem) The Hessian of a function $f : \mathbb{R}^n \to \mathbb{R}$ which has continuous second partial derivatives is always a symmetric matrix.

Proof. The $(i,j)$-th entry of $D^2f$ is $f_{x_ix_j}$, and by Clairaut’s Theorem this is the same as the $(j,i)$-th entry, $f_{x_jx_i}$. \hfill \Box

Example 1.8. Let’s revisit the example above, where we computed the Hessian of $f(x,y) = \ln(e^x + e^y)$. We found

$$D^2f = \frac{e^x e^y}{(e^x + e^y)^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

which is clearly symmetric!

Let’s see some examples of symmetric matrices.

Example 1.9. For $2 \times 2$ matrices, the condition that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is symmetric comes down to the condition that $b = c$. Hence, every $2 \times 2$ symmetric matrix can be written as

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

Example 1.10. For $3 \times 3$ matrices, the condition that $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ be symmetric is equivalent to the equalities $b = d$, $c = g$, and $f = h$. So every $3 \times 3$ symmetric matrix can be written as

$$A = \begin{bmatrix} a & b & c \\ b & e & f \\ c & f & i \end{bmatrix}$$

Example 1.11. Every (square) diagonal matrix, i.e. a matrix with non-zero entries only down the main diagonal (going from top left to bottom right), is automatically diagonal. Hence

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

is a diagonal matrix. However, it is clearly not true that every symmetric matrix is diagonal!

Despite this last remark, that not every symmetric matrix is diagonal, there is a good sense in which every symmetric matrix is almost diagonal. This is the content of the next very important theorem in linear algebra:

Theorem 1.12. (The Spectral Theorem for Symmetric Matrices) Let $A$ be an $n \times n$ symmetric matrix. Then there exist $n$ mutually orthogonal unit vectors $u_1, u_2, \ldots, u_n$ and $n$ real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ (which can possibly be zero) so that

$$Au_1 = \lambda_1 u_1$$
$$Au_2 = \lambda_2 u_2$$
$$\vdots$$
$$Au_n = \lambda_n u_n$$

Moreover the collection of $n$ real numbers $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ depend only on $A$ (and not on the particular choice of $u_1, \ldots, u_n$).
Proof. See HW 6, where you are asked to prove the Spectral Theorem (at least for $2 \times 2$) matrices using Calculus III! □

Why should you think of this condition of having $u_1, \ldots, u_n$ mutually orthogonal with $Au_i = \lambda_i u_i$ as “almost diagonal?” It is because diagonal matrices clearly have this property! If

$$A = \begin{bmatrix}
  a_1 & 0 & \cdots & 0 \\
  0 & a_2 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & a_n
\end{bmatrix}$$

is diagonal, then we can see that if we take $u_1 = e_1, \ldots, u_n = e_n$ and $\lambda_1 = a_1, \ldots, \lambda_n = a_n$, then the condition is clearly satisfied, since the standard basis $e_1, \ldots, e_n$ is clearly a collection of mutually orthogonal unit vectors and

$$Ae_i = a_i e_i$$

So in some sense, any symmetric matrix would be diagonal if we were allowed to “replace” the standard basis $e_1, \ldots, e_n$ by a new basis given by $u_1, \ldots, u_n$.

The following example is just to show a non-trivial example of the Spectral Theorem in action. You will not have to know the techniques used in the example for this class, but we include it for completeness.

**Example 1.13.** Find a pair of orthogonal unit vectors $u_1, u_2$ and the numbers $\lambda_1, \lambda_2$ so that $Au_i = \lambda_i u_i$ where

$$A = \begin{bmatrix}
  1 & -1 \\
  -1 & 1
\end{bmatrix}$$

We note that since the top row of $A$ is $-1$ times the bottom row of $A$ that something interesting should happen if we apply $A$ to the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. We compute

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1-1 \\ -1+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

So if we took $u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, this would satisfy $Au_1 = \lambda_1 u_1$ with $\lambda_1 = 0$. To find a $u_2$, we simply look for a vector orthogonal to $u_1$. Simple inspection tells us to try $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, so we compute

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

so if we take $u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, then we have $Au_2 = \lambda_2 u_2$ with $\lambda_2 = 2$.

The appeal of the Spectral Theorem for us is that it lets us define the following very simple invariant of symmetric matrices.

**Definition 1.14.** Let $A$ be a symmetric matrix, and suppose that $\det A \neq 0$ (this is equivalent to asking that none of the $\lambda_i$ are 0). Then we say $A$ has signature

$$(p, n-p)$$

if there are exactly $p$ positive $\lambda_i$ and $n-p$ negative $\lambda_i$.

**Example 1.15.** The matrix

$$A = \begin{bmatrix}
  2 & 0 & 0 \\
  0 & 5 & 0 \\
  0 & 0 & -1
\end{bmatrix}$$

has signature $(2, 1)$, since by the discussion above, $\lambda_1 = 2, \lambda_2 = 5, \lambda_3 = -1$, and two of these numbers are positive while one is negative.
Example 1.16. The matrix
\[
A = \begin{bmatrix}
1 & -1 \\
-1 & 1
\end{bmatrix}
\]
does not have a well-defined signature since we computed above that \(\lambda_1 = 0, \lambda_2 = 2\), and one of these two numbers is 0 (equivalently, we could notice that \(\det A = 0\)).

1.3. The Hessian and the Second Derivative Test. Equipped with some linear algebra, let’s come back to the second derivative test. The main object of interest was the quantity
\[
g_{u}''(0)
\]
where
\[
g_{u}(h) = f(a + hu)
\]
is the “cross section” of \(f\). To proceed, let’s find a formula for
\[
g_{u}'(h) = \frac{d}{dh}f(a + hu)
\]
We can compute the left hand side of this equation using the chain rule: function \(g_{u}\) is the composition
\[
x \mapsto f(x)
\]
\[
\mathbb{R} \rightarrow \mathbb{R}^n \xrightarrow{f} \mathbb{R}
\]
and so
\[
g_{u}'(h) = f_{x_1}(a + hu)u_1 + f_{x_2}(a + hu)u_2 + \ldots + f_{x_n}(a + hu)u_n
\]
But we are more interested in the second derivative \(g_{u}''(h)\), so we differentiate again to find
\[
g_{u}''(h) = \sum_{i=1}^{n} u_i \frac{d}{dh} f_{x_i}(a + hu)
\]
and repeating the same chain rule computation, but with \(f\) replaced by \(f_{x_i}\), i.e.,
\[
x \mapsto f_{x_i}(x)
\]
\[
\mathbb{R} \rightarrow \mathbb{R}^n \xrightarrow{f_{x_i}} \mathbb{R}
\]
we find the left hand side is
\[
g_{u}''(h) = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i f_{x_i x_j}(a + hu)u_j
\]
and evaluating at \(h = 0\), we find
\[
g_{u}''(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i f_{x_i x_j}(a)u_j
\]
Let’s take a second and process what this notation means for the case \(n = 2\). This is saying that
\[
g_{u}''(0) = \begin{bmatrix} u_1 f_{xx}(a) + u_1 f_{xy}(a) + u_2 f_{yx}(a) + u_2 f_{yy}(a) \\
                              u_1 f_{yx}(a) + u_2 f_{yy}(a)
\end{bmatrix} = \begin{bmatrix} u_1 \end{bmatrix} (D^2 f|_{x=a}) \begin{bmatrix} u_1 \end{bmatrix}
\]
In fact, it is easy to see that this is always true! That is,
\[
g_{u}''(0) = (D^2 f|_{x=a}) u
\]
Now, let’s apply the Spectral Theorem to the Hessian $D^2 f|_{x=a}$. This gives us $n$ mutually orthogonal unit vectors $u_1, \ldots, u_n$, and together with real numbers $\lambda_1, \ldots, \lambda_n$ so that $D^2 f|_{x=a} u_i = \lambda_i u_i$. Putting this in the expression for $g''_u(0)$ gives

$$g''_u(0) = \left( D^2 f|_{x=a} \right) u$$

We can say even more: since we can write any unit vector $u = \sum_{i=1}^n c_i u_i$, where $u \cdot u = \sum_{i=1}^n c_i^2 = 1$, we find that

$$g''_u(0) = \left( \sum_{i=1}^n c_i \lambda_i ight) u$$

Then we find that

$$g''_u(0) = \sum_{i=1}^n \lambda_i c_i^2$$

Note that this quantity is positive for any choice of $c_i$, i.e. for any $u$ if all of the $\lambda_i$ are positive, and is negative for all $u$ if all the $\lambda_i$ are negative. If some of the $\lambda_i$ are positive and some of the $\lambda_i$ are negative, then there are some $u$ for which this quantity is positive and some for which this is negative, but “how many” $u$ are positive and “how many” are negative is controlled by the number of $\lambda_i$ which are positive or negative. This gives us, combining with the primitive form of the second derivative test, the following theorem:

**Theorem 1.17.** (Second Derivative Test, refined form) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function with continuous second derivatives near a critical point $a$. If $\det D^2 f|_{x=a} \neq 0$, then if

$$\begin{cases} 
D^2 f|_{x=a} & \text{has signature (n,0) then } f \text{ has a local minimum at } a \\
D^2 f|_{x=a} & \text{has signature (0,n) then } f \text{ has a local maximum at } a \\
D^2 f|_{x=a} & \text{has signature (p, n-p) then } f \text{ has a saddle point at } a \\
\end{cases}$$

If $\det D^2 f|_{x=a} = 0$ then the test is inconclusive.

The term “$(p, n-p)$ saddle point” simply means that there are $p$ independent directions where the graph of the function curves downward, and $n-p$ independent directions where it curves up.

Let’s see the typical examples for this test.

**Example 1.18.** Let $f(x, y) = x^2 + y^2$. (Draw picture of graph.) Then there is only one critical point, and it is at $(0,0)$. If we compute the Hessian, we get

$$D^2 f = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

So $\lambda_1 = \lambda_2 = 2$, and thus the signature of the Hessian at $0$ is $(2,0)$. This function has a minimum at $0$.

**Example 1.19.** Let $f(x, y) = -x^2 - y^2$. (Draw picture of graph.) Then there is only one critical point, and it is at $(0,0)$. If we compute the Hessian, we get

$$D^2 f = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$$

So $\lambda_1 = \lambda_2 = -2$, and thus the signature of the Hessian at $0$ is $(0,-2)$. This function has a maximum at $0$.

**Example 1.20.** Let $f(x, y) = x^2 + y$. (Draw picture of graph.) Then there is only one critical point, and it is at $(0,0)$. If we compute the Hessian, we get

$$D^2 f = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

So $\lambda_1 = 2, \lambda_2 = -2$, and thus the signature of the Hessian at $0$ is $(1,1)$. This function has a saddle point at $0$; one direction goes up, the other direction goes down.

OK, enough about the second derivative test. Let’s do something different (and more useful!).


2. INTRODUCTION TO CONSTRAINED OPTIMIZATION

Let’s start with a simple example.

Example 2.1. Maximize the area enclosed by a rectangular fence when you only have 4 meters of fence to work with.

Set up problem: if call the two side lengths \( x \) and \( y \), we want to maximize \( f(x, y) = xy \) subject to the restriction \( 2x + 2y = 4 \). We use this constraint and write \( f \) as a function of only \( x \), to find that we are maximizing the function of one variable \( x(2 - x) \). We do so, and find \( x = 1, y = 1 \) is the maximum.

This is the sort of typical problem Lagrange multipliers deals with. We want to maximize a function of many variables, say two variables, \( f(x, y) \) but subject to a constraint \( g(x, y) = c \). So in the case above \( f(x, y) = xy, g(x, y) = 2x+2y, c = 4 \). However, this case was quite exceptional! We could easily parameterize the constraint \( g(x, y) = c \) (it is just a line!) and essentially this is what we did, calling our parameter \( x \), and then restricting \( f \) to the constraint using the parameterization.

That is, there is the following semi-general approach to these problems. We have the following set up.

\[
\begin{align*}
\mathbb{R} & \xrightarrow{r} \mathbb{R}^2 & \xrightarrow{f} & \mathbb{R} \\
 t & \mapsto r(t) & & f(x) \\
 x & \mapsto f(x)
\end{align*}
\]

and then solving the unconstrained maximization problem for the function \( f(r(t)) \).

Why do I call this technique only “semi-general?” It is because sometimes it can be very hard to parameterize the constraint! Moreover this method can be a little painful if there is a parameterization, but it’s ugly.

What we would like is a technique to solve these optimization problems with constraints that does not need to do the extra step of parameterizing the constraint.

OK, so let’s think about it. For simplicity, let’s just work with a function of two variables for now. So suppose we want to maximize \( f \) subject to a constraint, which is a level curve \( g(x, y) = c \). (Draw a picture of \( g(x, y) = c \).) Now, we know that the gradient of \( f \), \( \nabla f(x) \) is a vector based at \( x \) pointing in the direction of greatest increase for \( f \). So if we wanted to maximize \( f \) on the curve \( g(x, y) \) it makes sense to think about what the gradient of \( f \) is doing on that curve. (Draw picture of the gradient.) If there is any component of the gradient pointing tangential to the curve, then we can do better, i.e. we can increase the value of \( f \) by moving in that direction along the curve. So we would keep moving until the gradient became orthogonal to the curve itself.

However, what does that mean? The curve has a unique normal vector at a point \( x \), and it is given by \( \nabla g(x) \). So asking that \( \nabla f(x) \) be orthogonal to the curve is asking that is be parallel to \( \nabla g(x) \), i.e. it is asking that there exist a \( \lambda \) so that

\[
\nabla f(x) = \lambda \nabla g(x)
\]

Of course, any such point \( x \) must also satisfy the constraint equation

\[
g(x) = c
\]

This gives rise to a system of three equations and three unknowns \( x, y, \lambda \), namely

\[
\frac{\partial f}{\partial x}(x, y) = \lambda \frac{\partial g}{\partial x}(x, y) \\
\frac{\partial f}{\partial y}(x, y) = \lambda \frac{\partial g}{\partial y}(x, y) \\
g(x, y) = c
\]

Let’s see how this method applies to our earlier example.

Example 2.2. Maximize the area enclosed by a rectangular fence when you only have 4 meters of fence to work with.
We take \( f(x, y) = xy \), \( g(x, y) = 2x + 2y \), and \( g(x, y) = 4 \) to be the constraint. We compute
\[
\nabla f = \begin{bmatrix} y \\ x \end{bmatrix}, \quad \nabla g = \begin{bmatrix} 2 \\ 2 \end{bmatrix}
\
\]
So we get the equations
\[
y = 2\lambda \\
x = 2\lambda \\
4 = 2x + 2y
\]
and solving we see \( x = y \) by the first two equations, so by the last equation \( x = y = 1 \).

This is the method of Lagrange multipliers. It works for functions from \( \mathbb{R}^n \to \mathbb{R} \); the proof is the same. There is another, similar but slightly different proof in the book using level curves. (Present proof if time.)

**Theorem 2.3.** Suppose that \( f \) and \( g \) are differentiable functions \( \mathbb{R}^n \to \mathbb{R} \). Then any maximum or minimum of \( f(x) \) subject to the constraint \( g(x) = c \) must satisfy the equations
\[
\nabla f(x) = \lambda \nabla g(x) \\
g(x) = c
\]
Note that this is really a system of \( n + 1 \) equations and \( n + 1 \) unknowns.

Let’s see another example in action.

**Example 2.4.** A rectangular box without a lid is being made from 12 square meters of cardboard. Find the maximum volume of the box. (Answer: \( z = 1 \), \( z = 2 \), \( y = 2 \).)

We set up the problem: we have a box with dimensions \( x, y, z \). The volume, which we are trying to maximize, is given by
\[
f(x, y, z) = xyz
\]
while the constraint is that we only have 12 square meters of cardboard, i.e.
\[
xy + 2xz + 2yz = 12
\]
calling \( g(x, y, z) = xy + 2xz + 2yz \), this constraint is that \( g(x, y, z) = 12 \).

So now we do the method of Lagrange multipliers. This means we compute
\[
\nabla f = \begin{bmatrix} yz \\ xz \\ xy \end{bmatrix}
\]
and
\[
\nabla g = \begin{bmatrix} y + 2z \\ x + 2z \\ 2x + 2y \end{bmatrix}
\]
so the Lagrange multipliers method gives us the following 4 equations with 4 unknowns:
\[
yz = \lambda(y + 2z) \\
xz = \lambda(x + 2z) \\
xy = \lambda(2x + 2y) \\
12 = xy + 2xz + 2yz
\]
Note that if we multiply the first equation by \( x \), the second equation by \( y \), and the third equation by \( z \), then we get the same left hand sides \( xyz \). This means we get equations
\[
\lambda(xy + 2xz) = \lambda(2xz + 2yz)
\]
and
\[
\lambda(xy + 2yz) = \lambda(2xz + 2yz)
\]
Now, \( \lambda \neq 0 \) since that would violate the constraint, so we can cancel it. We can also cancel other terms in these two equations to find
\[
xz = 2yz \\
xy = 2xz
\]
We also note that none of $x, y, z$ can be 0, since then the box would have volume 0, which is clearly not the maximum. So we find

\[
x = 2z \\
y = 2z
\]

So, plugging into the constraint, we find

\[12z^2 = 12\]

so $z = 1$ and $x, y = 2(1) = 2$. 