Monday, October 12 was Thanksgiving Holiday

Lecture 13

Optimization problems with constraints - the method of Lagrange multipliers

(Relevant section from the textbook by Stewart: 14.8)

In Lecture 11, we considered an optimization problem with constraints. The problem was solved by using the constraint to express one variable in terms of the other, hence reducing the dimensionality of the problem. Sometimes, it is not convenient, or even possible, to perform such a reduction of dimensionality. In this lecture, we’ll examine a general method to solve optimization problems with constraints, the so-called method of Lagrange multipliers. We first consider another optimization problem with constraints as a motivation.

Example: Find the point \( P \) on the plane \( x + y - 2z = 6 \) which lies closest to the origin.

The plane \( x + y - 2z = 6 \) passes through the three points \((6,0,0)\), \((0,6,0)\) and \((0,0,-3)\) on, respectively, the \( x \), \( y \) and \( z \)-axes. As such, it forms a tetrahedron with the \( xy \), \( yz \) and \( xz \)-planes that lies below the \( xy \)-plane. We expect the desired point to lie on the part of this plane that forms a triangular face of this tetrahedron.

The distance of a point \( P \) with coordinate \((x, y, z)\) to the origin is given by \( D = \sqrt{x^2 + y^2 + z^2} \). We are asked to minimize \( D \). This is equivalent to minimizing the square of the distance

\[
D^2 = x^2 + y^2 + z^2 = f(x, y, z),
\]

which is an easier problem since there are no square roots.

But there is a restriction: the points \((x, y, z)\) must lie on the plane \( x + y - 2z = 6 \). This is an example of a minimization problem with constraints.

Most real-world optimization problems involve constraints. There are often restrictions on the values that can be assumed by the variables concerned. For example, the amount of money in one’s portfolio or the concentration of a chemical species must be non-negative. In addition, in many problems, there are conservation rules, e.g., conservation of mass. In some chemical reactions, for examples, the sum of the concentrations of particular chemical species must be constant, so that no mass is created or destroyed.
We now come to a fundamental point regarding problems with constraints: Each constraint in a problem that can be expressed in the form

$$F(x, y, z) = 0$$

reduces the *dimensionality* of the problem, i.e., the number of independent variables, by one. In the problem posed above, there are three variables, $x, y$ and $z$ needed to uniquely address a point $P$ in $\mathbb{R}^3$. But the condition that the point must lie on the given plane can be written as the constraint

$$F(x, y, z) = x + y - 2z - 6 = 0.$$  

Therefore the number of independent variables in this problem is two. We shall rewrite this problem as follows:

Minimize $$f(x, y, z) = x^2 + y^2 + z^2$$ subject to $$x + y - 2z - 6 = 0.$$  

In this problem, as in the one considered in Lecture 11, we can use the constraint to express one variable in terms of the other two. For example, let us consider $z$ as a function of $x$ and $y$, i.e., $z(x, y)$:

$$z = \frac{1}{2}x + \frac{1}{2}y - 3.$$  

This produces a function of two variables $h(x, y)$ that we must now minimize over $\mathbb{R}^2$, i.e.,

$$D^2 = h(x, y) = x^2 + y^2 + \left(\frac{1}{2}x + \frac{1}{2}y - 3\right)^2, \quad (x, y) \in \mathbb{R}^2.$$  

**Note:** The following details were not presented in class, in order to save some time

We now determine the critical points of $h(x, y)$:

$$\frac{\partial h}{\partial x} = 2x + 2 \left(\frac{1}{2}x + \frac{1}{2}y - 3\right) \cdot \frac{1}{2}$$

$$\frac{\partial h}{\partial y} = 2y + 2 \left(\frac{1}{2}x + \frac{1}{2}y - 3\right) \cdot \frac{1}{2}$$

The condition for a critical point $\nabla h(x, y) = (0, 0)$ leads to the equations

$$\frac{5}{2}x + \frac{1}{2}y = 3$$

$$\frac{1}{2}x + \frac{5}{2}y = 3.$$
This system has the unique solution \( x = y = 1 \).

We now use the constraint to obtain \( z \):

\[
z = \frac{1}{2} x + \frac{1}{2} y - 3 = -2.
\]

Therefore the desired point \( P \) is \((1, 1, -2)\). The distance between \( P \) and the origin is \( \sqrt{1 + 1 + 4} = \sqrt{6} \).

We should really check that the above point represents the absolute minimum. Firstly, the point \((x, y) = (1, 1)\) was found to be the only critical point of \( h(x, y) \) defined above. A look at the second-order derivatives at \( P \) gives

\[
A = \frac{\partial^2 h}{\partial x^2} = \frac{5}{2}, \quad B = \frac{\partial^2 h}{\partial x \partial y} = \frac{1}{2}, \quad C = \frac{\partial^2 h}{\partial y^2} = \frac{5}{2},
\]

so that \( B^2 - AC = -6 < 0 \). Since \( A > 0 \), we have that \((1, 1)\) is a relative minimum. Since the function \( h(x, y) \) cannot be negative and can assume arbitrarily large values, we can conclude that \((1, 1)\) is an absolute minimum.

The method of Lagrange multipliers

The basic problem of optimization with a constraint can be formulated as follows:

\[
\text{minimize or maximize } f(x, y, z) \\
\text{subject to the constraint } F(x, y, z) = 0.
\]

(We can add more constraints, e.g., \( G(x, y, z) \), and will do so later.)

Here’s the final result – we’ll provide a brief derivation later:

1. First construct the so-called “Lagrangian function”

\[
L(x, y, z, \lambda) = f(x, y, z) + \lambda F(x, y, z),
\]

where “\( \lambda \)” is known as a “Lagrange multiplier.”

2. Now minimize or maximize the function \( L(x, y, z, \lambda) \) with respect to the four variables \((x, y, z, \lambda) \in \mathbb{R}^4 \).

You might be saying to yourself, “Wow! It doesn’t look like we’ve simplified the problem. We now have four variables over which to optimize. We actually went down from three to two variables when we used the constraint to remove a variable. What gives?”
The answer is that the method of Lagrange multipliers is a general method that is effective in solving a wide variety of problems. It may not always be possible to express one variable in terms of the others (recall our discussion on implicit functions). Furthermore, the method of Lagrangians is very useful in more general or abstract problems involving an arbitrary number of independent variables and/or constraints. For example, in a future course or courses in Physics (e.g., thermal physics, statistical mechanics), you should see a derivation of the famous “Boltzman distribution” of the energies of atoms in an ideal gas using Lagrange multipliers.

**Example:** Let us return to the optimization problem with constraints discussed earlier: Find the point $P$ on the plane $x + y - 2z = 6$ that lies closest to the origin. Recall that we sought to minimize the square of the distance:

$$\text{Minimize } f(x, y, z) = x^2 + y^2 + z^2$$

subject to

$$x + y - 2z - 6 = 0.$$  

**Solution:** The Lagrangian function associated with this problem is

$$L(x, y, z, \lambda) = f(x, y, z) + \lambda F(x, y, z)$$

$$= x^2 + y^2 + z^2 + \lambda(x + y - 2z - 6).$$

We must find the critical points of $L$ in terms of the four variables $x, y, z$ and $\lambda$:

$$\frac{\partial L}{\partial x} = 2x + \lambda = 0$$

$$\frac{\partial L}{\partial y} = 2y + \lambda = 0$$

$$\frac{\partial L}{\partial z} = 2z - 2\lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = x + y - 2z - 6 = 0.$$ 

Note that the final equation simply corresponds to the constraint applied to the problem. Clever, eh?

There are often several ways to solve problems involving Lagrangians and Lagrangian multipliers. The most important point to remember is that one method does not often work for all problems. In this case, we can find the critical point rather easily as follows. We use the first three equations to express $x, y$ and $z$ in terms of $\lambda$:

$$x = \frac{-\lambda}{2}, \quad y = \frac{-\lambda}{2}, \quad z = \lambda.$$  

94
We now substitute these results into the fourth equation:

\[-\frac{\lambda}{2} - \frac{\lambda}{2} - 2\lambda - 6 = 0 \implies 3\lambda = -6,\]

which implies that \(\lambda = -2\). From the above three equations, we have determined \(x, y\) and \(z\):

\[x = 1, \quad y = 1, \quad z = -2.\]  

(15)

Therefore the desired point is \((1, 1, -2)\), which is in agreement with the result obtained in the previous lecture.

We continue with another illustrative example.

**Example:**

Maximize/minimize \(f(x, y, z) = xyz\) on the ellipse \(x^2 + 2y^2 + 3z^2 = 1\).  

(16)

The ellipse represents the constraint in this problem. We first express this constraint in the form \(F(x, y, z) = 0\), i.e.,

\[F(x, y, z) = x^2 + 2y^2 + 3z^2 - 1 = 0.\]

(17)

The Lagrangian associated with this problem is then

\[L(x, y, z, \lambda) = xyz + \lambda(x^2 + 2y^2 + 3z^2 - 1).\]

(18)

The critical points of the Lagrangian must satisfy the following equations

\[
\begin{align*}
\frac{\partial L}{\partial x} &= yz + 2\lambda x = 0 \quad \text{(a)} \\
\frac{\partial L}{\partial y} &= xz + 4\lambda y = 0 \quad \text{(b)} \\
\frac{\partial L}{\partial z} &= xy + 6\lambda z = 0 \quad \text{(c)}
\end{align*}
\]

The final condition \(\frac{\partial L}{\partial \lambda} = 0\) yields the constraint.

Once again, we’re faced with the problem of solving this system of equations, which is now nonlinear. Here is a “trick” that works because of the symmetry of the problem. (It won’t always
Multiply the first equation by \( x \), the second by \( y \) and the third by \( z \):

\[
\begin{align*}
xyz + 2\lambda x^2 &= 0 \quad (d) \\
xyz + 4\lambda y^2 &= 0 \quad (e) \\
xyz + 6\lambda z^2 &= 0 \quad (f)
\end{align*}
\]

There are a number of possible paths to pursue, but we consider the following: Since there is an \( xyz \) term in each equation, we can equate the other terms, i.e.,

\[
-2\lambda x^2 = -4\lambda y^2 = -6\lambda z^2,
\]

and then remove the minus signs. Or we can subtract (e) from (d),

\[
2\lambda x^2 - 4\lambda y^2 = 0, \tag{22}
\]

then (f) from (e), and finally (f) from (d). The final result is the same:

\[
2\lambda x^2 = 4\lambda y^2 = 6\lambda z^2, \tag{23}
\]

or

\[
\lambda x^2 = 2\lambda y^2 = 3\lambda z^2. \tag{24}
\]

It is tempting to divide out the \( \lambda \), but we must consider the possibility that \( \lambda = 0 \).

**Case 1:** \( \lambda = 0 \) It follows, from (a), (b) and (c), that at least two of \( x, y \) and \( z \) are zero, implying that \( f(x, y, z) = 0 \). We can easily solve for these points, since they lie on the ellipse:

\[
(\pm 1, 0, 0), \quad \left(0, \pm \frac{1}{\sqrt{2}}, 0\right), \quad \left(0, 0, \pm \frac{1}{\sqrt{3}}, \right). \tag{25}
\]

They lie at the extreme top and bottom and sides of the ellipse. At all of these six points, \( f(x, y, z) = 0 \).

**Case 2:** \( \lambda \neq 0 \) In this case, Eq. (24) becomes

\[
x^2 = 2y^2 = 3z^2. \tag{26}
\]

The point \((0, 0, 0)\) is inadmissible, since it does not lie on the ellipse. Since this point must lie on the ellipse, we set

\[
x^2 = 2y^2 = 3z^2 = t. \tag{27}
\]
Substitution into the equation for the ellipse yields,

\[ x^2 + 2y^2 + 3z^2 = 3t = 1, \quad \text{implying that} \quad t = \frac{1}{3}. \quad (28) \]

From Eq. (27), we have

\[ x = \pm \frac{1}{\sqrt{3}}, \quad y = \pm \frac{1}{\sqrt{6}}, \quad z = \pm \frac{1}{3}. \quad (29) \]

In summary, we have determined \( 2^3 = 8 \) points that lie on the ellipse which are contenders for maximum and minimum points of \( f(x, y, z) = xyz \). It is clear that if the two values of \( f \) produced by these points are

\[ f_{\min} = -\frac{1}{9\sqrt{2}}, \quad f_{\max} = \frac{1}{9\sqrt{2}}. \quad (30) \]

**An alternate solution (it was not presented in class)**

Here we present another method of solving the above problem. Admittedly, it is a longer method. But it is based on another idea that may be useful in other situations. It’s always helpful to be able to consider more than one method for the solution of a problem.

We start again with Eqs. (d), (e) and (f) above. Instead of removing the \( xyz \) terms, we add up both sides of the three equations:

\[ 3xyz + 2\lambda(x^2 + 2y^2 + 3z^2) = 0. \quad (31) \]

Because of the constraint, this equation reduces to

\[ 3xyz + 2\lambda = 0 \quad \text{or} \quad 3xyz = -2\lambda \quad (g) \quad (32) \]

From Eq. (a), multiplying both sides by \( x \):

\[ 3xyz = -6\lambda x^2. \quad (33) \]

Using this result with along (g), we have

\[ -6\lambda x^2 = -2\lambda. \quad (34) \]

There are two possibilities:

1. \( \lambda = 0 \) or
2. \( x^2 = \frac{1}{3} \) which implies that \( x = \pm \frac{1}{\sqrt{3}} \).
Now multiply both sides of (b) by $y$:

$$3xyz = -12\lambda y^2.$$  \hfill (35)

Using this result along with (g) gives

$$-12\lambda y^2 = -2\lambda.$$  \hfill (36)

There are two possibilities:

1. $\lambda = 0$ or
2. $y^2 = \frac{1}{6}$ which implies that $y = \pm \frac{1}{\sqrt{6}}$.

Finally, multiply both sides of (c) by $z$:

$$3xyz = -18\lambda y^2.$$  \hfill (37)

Using this result along with (g) gives

$$-18\lambda y^2 = -2\lambda.$$  \hfill (38)

There are two possibilities:

1. $\lambda = 0$ or
2. $z^2 = \frac{1}{9}$ which implies that $z = \pm \frac{1}{3}$.

We see that the case $\lambda \neq 0$ yields the same points that were found by the previous method. And the case $\lambda = 0$ can be treated in the same way as before. This illustrates that there may be more than one way to solve a problem involving Lagrange multipliers.

Note that in both of the examples examined above, we did not bother with the task of determining whether or not the critical points of the Lagrangian $L$, hence the function $f$, were local minima, maxima or saddle points. (In fact, we really couldn’t do this because we haven’t covered the second derivative test for functions of more than two variables!) Just as in the earlier discussion of finding absolute maxima and minima of functions on restricted domains, it is sufficient to find critical points and evaluate the functions at these critical points, from which we can determine the absolute maximum and/or minimum values. That being said, it may be necessary, in some cases, to examine the behaviour of a function for arbitrarily large values of the independent variables $x$, $y$, etc., to ensure that we have, in fact, found an absolute maximum or minimum (or both).
Lecture 14

A brief look at the theory behind the method of Lagrange multipliers

(Relevant section from the textbook by Stewart: 14.8)

We now examine the Lagrange multiplier method more closely. For simplicity, we consider only functions of two variables and examine the optimization problem

maximize/minimize $f(x, y)$
subject to $F(x, y) = 0$.

The constraint $F(x, y) = 0$ defines a curve $C$ in $\mathbb{R}^2$ as sketched in the figure below. The points $(x, y)$ that lie on $C$ represent the only allowable values of $(x, y)$ that may be considered in the evaluation of $f(x, y)$. (For example, recall the path of the ant on the hotplate.) Associated with each point $P(x, y)$ on this curve $C$ is the point $Q(x, y, f(x, y))$ that lies on the graph of $f(x, y)$, in other words, the surface $z = f(x, y)$. As the point $P$ moves along the $xy$-plane, it traces out a curve $D$ in $\mathbb{R}^3$ which lies on the surface $z = f(x, y)$.

We are interested in finding maximum or minimum values of $f(x, y)$ evaluated over the curve $C$. At the points where such extrema can occur, the rate of change of $f(x, y)$ must be zero, but in the direction of the curve $C$ at such points. But the direction of the curve at a point $P$ on the curve is given by the tangent to the curve at $P$. Therefore, it follows that the condition for a maximum or minimum of $f(x, y)$ is that the directional derivative of $f(x, y)$ in the direction of the tangent vector to curve $C$ is zero.

To see this a little more clearly, let us assume that we can parametrize the curve $C$ as $(x(t), y(t))$. As we move over the curve $C$, we consider the value of $f(x(t), y(t))$ evaluated at points on $C$. We now
look for points at which the rate of change of $f(x(t), y(t))$ is zero, in other words, the total derivative
\[
\frac{df}{dt} = 0.
\] (39)
But the total derivative of $f$ with respect to $t$ is given by
\[
\frac{d}{dt} f(x(t), y(t)) = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} = 0.
\] (40)
But this equation can be rewritten as
\[
\vec{\nabla} f \cdot \mathbf{v} = 0.
\] (41)
This again verifies that the directional derivative of $f$ in the direction of $\mathbf{v}$ is zero. We shall return to this result.

Now recall that we are moving along the curve $C (x(t), y(t))$ because of the constraint $F(x, y) = 0$. This means that
\[
F(x(t), y(t)) = 0, \quad \text{for all } t.
\] (42)
In other words, the value of $F(x, y)$ is constant for all $t$: After all, curve $C$ is a zero-level set for the function $F$! This means that
\[
\frac{dF}{dt} = 0.
\] (43)
But the total derivative of $F$ with respect to $t$ is given by
\[
\frac{d}{dt} F(x(t), y(t)) = \frac{\partial F}{\partial x} \frac{dx}{dt} + \frac{\partial F}{\partial y} \frac{dy}{dt} = 0.
\] (44)
But this equation can be rewritten as
\[
\vec{\nabla} F \cdot \mathbf{v} = 0, \quad \text{for all } t.
\] (45)
Eqs. (41) and (45) imply that the two vectors $\vec{\nabla} f$ and $\vec{\nabla} F$ are perpendicular to the tangent vector $\mathbf{v}$ at a relative maximum or minimum of $f$, as measured over curve $C$. The only way that two vectors $\mathbf{a}$ and $\mathbf{b}$ can be perpendicular to a given vector $\mathbf{u}$ in the plane is that one is a multiple of the other, i.e., $\mathbf{b} = K \mathbf{a}$ for some constant $K$. Therefore, at the critical point of $f$ as measured over $C$, we have the result that
\[
\vec{\nabla} f = K \vec{\nabla} F.
\] (46)
We may rewrite Eq. (46) as
\[
\vec{\nabla} [f(x, y) - K F(x, y)] = 0,
\] (47)
The expression in brackets is the Lagrangian function associated with this optimization problem with constraint, with $\lambda = -K$:

$$L(x, y, \lambda) = f(x, y) + \lambda F(x, y).$$  \hfill (48)

In summary, the condition for a relative maximum or minimum of $f(x, y)$ as evaluated over the curve defined by the constraint $F(x, y) = 0$ is that the point be a critical point of the Lagrangian function $L(x, y, \lambda)$ associated with the optimization problem, i.e.

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = 0.$$  \hfill (49)

Note that the additional condition for a critical point in terms of $\lambda$,

$$\frac{\partial L}{\partial \lambda} = F(x, y) = 0,$$  \hfill (50)

which is the original constraint.

**Note:** You will probably see that Eq. (46) is how Stewart defines the method of Lagrange multipliers in his textbook. This way of approaching the problem is perfectly fine – one is simply bypassing the formal construction of the Lagrangian function $L$ and then searching for critical points. We have included the formal definition of $L$ in this course because it represents the more “traditional” way of discussing the subject, especially in Science and Engineering-oriented books. It may also be an easier approach to remember.

**An analysis in terms of level sets of $f$**

The above discussion may still seem somewhat mysterious. A look at how the level sets of the function $f(x, y)$ being optimized, and their relation to the constraint curve $C$, will help. (This examination was actually presented first in the lecture, before the above analysis.)

Let’s consider the following scenario, as sketched graphically below. Some level sets of the function $f(x, y)$ are drawn, along with the curve $C$ that is produced by the constraint $F(x, y) = 0$. As an observer travels along $C$, the values of $f$ measured by the observer will be given by the values of the level sets that the observer intersects along the journey.

Note that there is one special level set of $f$ that intersects the curve $C$ at only one point $P$, namely, the level set corresponding to $f(x, y) = 35$. As one approaches this point of intersection from one side of $P$, the observed value of $f$ increases until it attains the value 35 at $P$. As we continue,
therefore moving away from \( P \), the observed value of \( f \) then decreases. As a result, we have located a local maximum point of \( f \) subject to the constraint: It is the point \( P \).

There is something special about this point of intersection, \( P \): You’ll see that the two curves have a common tangent vector. If they didn’t, then the curve \( C \) would cross the level set of \( f \) at 35 and continue to encounter higher values of \( f \), in the same way that it crosses the level set of \( f \) at 30 and then proceeds to move on, finding larger values of \( f \).

If these two curves have a common tangent vector at point \( P \), it follows that their normal vectors are either parallel (point in same direction) or antiparallel (point in opposite direction). But what are these normal vectors?

1. For the function \( f \): We know that \( \pm \vec{\nabla} f \) (and multiples) are normal vectors to a level curve/set.

2. For the function \( F \): We are looking for the normal vector to the curve \( C \) defined by \( F(x, y) = 0 \).

   But, as mentioned earlier, this is the zero-level set of \( F \). Therefore \( \pm \vec{\nabla} F \) (and multiples) are normal vectors.

As a result, we may conclude that

\[
\vec{\nabla} f = K \vec{\nabla} F
\]  

at the critical point \( P \), for some constant \( K \in \mathbb{R} \). But this is the same result as in Eq. (46)!

An important note: Note that the gradient vector field of \( f \), namely \( \vec{\nabla} f \), is fixed in space, as is the gradient vector field of \( F \) over the curve \( C \). It is at these special points of intersection, where Eq. (51) is satisfied, however, where local max/min of \( f \) subject to the constraint are located.
Extension to higher dimensions

The above discussion can be extended to treat optimization problems for functions of more variables, e.g., \( f(x, y, z) \). In higher dimensions, however, the situation is a little more complicated – for example, if two vectors are perpendicular to a given vector \( \mathbf{v} \), it doesn’t follow that one is a multiple of the other. In \( \mathbb{R}^3 \), they must lie on the plane that has \( \mathbf{v} \) as its normal. In any case, the above discussion will hopefully give some idea as to why the Lagrangian method works.

The method of Lagrange multipliers for more than one constraint

The method of Lagrange multipliers can also accommodate several constraints. Each constraint will require a Lagrangian multiplier. For example, the Lagrangian associated with the optimization problem with two constraints

\[
\text{maximize/minimize } f(x, y, z) \\
\text{subject to } F(x, y, z) = 0 \text{ and } G(x, y, z) = 0,
\]

is given by

\[
L(x, y, z, \lambda, \mu) = f(x, y, z) + \lambda F(x, y, z) + \mu G(x, y, z). \tag{52}
\]

The necessary condition for a relative maximum or minimum is that

\[
\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial z} = 0, \quad \frac{\partial L}{\partial \lambda} = 0, \quad \frac{\partial L}{\partial \mu} = 0. \tag{53}
\]

Example: Find the minimum value of the function

\[
f(x, y, z) = x^2 + 2y^2 + z^2 \tag{54}
\]

subject to the constraints

\[
x + 2y + 3z = 1 \tag{55}
\]

\[
x - 2y + z = 5.
\]
Before we outline the solution of this problem, let us step back for a moment and consider its geometrical interpretation. The level sets of \( f(x, y, z) \) are concentric ellipsoids centered at the origin \((0, 0, 0)\). As we move outward, the values associated with these level sets increase.

The two constraints represent equations of planes. The fact that they must be satisfied simultaneously means that the planes are intersecting – in this case, their intersection produces a line in \( \mathbb{R}^3 \). As one moves along this line, closer and closer to the origin, the value of \( f \) on the line will decrease as it intersects level sets associated with lower and lower values of \( f \). If the line actually were to go through the origin (which is not the case, since \((0, 0, 0)\) does not satisfy any of the two equations), the value of \( f \) evaluated on the line would go to zero.) At some point, a minimal value of \( f \) will be attained, and the values will begin to increase.

**Solution:** The two constraints can be written in the form

\[
F(x, y, z) = x + 2y + 3z - 1 = 0 \quad (56)
\]
\[
G(x, y, z) = x - 2y + z - 5 = 0.
\]

The associated Lagrangian function then has the form

\[
L(x, y, z, \lambda, \mu) = x^2 + 2y^2 + z^2 + \lambda(x + 2y + 3z - 1) + \mu(x - 2y + z - 5). \quad (57)
\]

The conditions for a critical point become:

\[
\frac{\partial L}{\partial x} = 0 : \quad 2x + \lambda + \mu = 0 \quad (58)
\]
\[
\frac{\partial L}{\partial y} = 0 : \quad 4y + 2\lambda - 2\mu = 0
\]
\[
\frac{\partial L}{\partial z} = 0 : \quad 2z + 3\lambda + \mu = 0
\]
\[
\frac{\partial L}{\partial \lambda} = 0 : \quad x + 2y + 3z - 1 = 0
\]
\[
\frac{\partial L}{\partial \mu} = 0 : \quad x - 2y + z - 5 = 0.
\]

We search for \( x, y, z, \lambda \) and \( \mu \) that simultaneously satisfy these equations.

As mentioned in class, there are usually several ways to solve these equations. And a method that solves one problem will not necessarily apply to another. To solve this problem, one method is to use
the first three equations from above to express \( x, y \) and \( z \) in terms of \( \lambda \) and \( \mu \):

\[
\begin{align*}
x &= -\frac{1}{2}(\lambda + \mu), \\
y &= -\frac{1}{4}(2\lambda - 2\mu), \\
z &= -\frac{1}{2}(3\lambda + \mu).
\end{align*}
\]  

(59)

Now substitute these results into the final two equations, which represent the constraints:

\[
\begin{align*}
-6\lambda - \mu &= 1 \\
-\lambda - 2\mu &= 5.
\end{align*}
\]  

(60)

The solution of this simultaneous linear system is given by

\[
\lambda = \frac{3}{11}, \quad \mu = -\frac{29}{11}.
\]  

(61)

From these values, we compute \( x, y \) and \( z \) to be

\[
\begin{align*}
x &= \frac{13}{11}, \\
y &= -\frac{16}{11}, \\
z &= \frac{10}{11}.
\end{align*}
\]  

(62)

This is the only critical point for this problem. At this point, \( f(x, y, z) = \frac{71}{11} \). This must correspond to a global minimum since \( f(x, y, z) \) can assume arbitrary large values by letting \( x, y \) and \( z \) become arbitrarily large while they satisfy the two constraints.

**An important application of Lagrange multipliers to Physics – the Boltzmann distribution of Statistical Mechanics**

The discussion in this section is intended to be brief. We outline the application of method of Lagrangian multipliers a fundamental problem in Statistical Mechanics: finding the most probable distribution of energies assumed by a system of atoms or molecules. From now on, we simply refer to these particles as molecules.

Consider a system of \( N \) independent, identical and distinguishable atoms or molecules, for example, a container of oxygen gas. By “distinguishable,” we mean that we can index each molecule uniquely and keep track of it. We assume that each molecule can exist in one of \( n \) states, \( 1, 2, \cdots, n \), with respective energies \( E_1, E_2, \cdots, E_n \). (Examples of these energies: the electronic energies that can be assumed by an atom, the vibrational energies of a diatomic molecule.) We'll also let \( N_k \) denote the “occupation number” of the \( k \)th state, i.e. the number of molecules in that state.
We have already arrived at our first constraint, namely,

$$N_1 + N_2 + \cdots + N_n = N.$$  \hfill (63)

We also impose the condition that the total energy of the system is constant, i.e.,

$$N_1 E_1 + N_2 E_2 + \cdots + N_n E_n = E.$$  \hfill (64)

The first question to be considered is the following:

How many different ways are there of arranging \( N \) different molecules among these different states?

This is a combinatorial problem with which you may be familiar. The answer is

$$W(N_1, N_2, \ldots, N_n) = \frac{N!}{N_1! N_2! \cdots N_n!}.$$  \hfill (65)

As an example, we consider the very simple case, \( N = 2, n = 2 \), i.e., two molecules and two energy states. There are four ways to arrange the molecules:

This is in agreement with the above formula, since the numbers of ways \( W(N_1N_2) \) corresponding to all possible \( (N_1, N_2) \) values are

$$W(2, 0) = \frac{2!}{0!2!} = 1, \quad W(0, 2) = \frac{2!}{2!0!} = 1, \quad W(1, 1) = \frac{2!}{1!1!} = 2.$$  \hfill (66)

The claim of Statistical Mechanics is that the “equilibrium” distribution of energies corresponds to the set of occupation numbers \( (N_1, N_2, \ldots, N_n) \) such that the function \( W(N_1, N_2, \ldots, N_n) \) is maximized, subject to the constraints in (63) and (64). This problem seems to be perfectly suited for the method of Lagrange multipliers.

There is just one minor technicality, however. The quantities \( N_k \) are integers, whereas all of our discussions of multivariable functions and their optimization is based on the assumption that the independent variables are continuous real variables.

For very large \( N \), as is the case for a container of gas molecules (\( N \) is on the order of Avogadro’s number, roughly \( 10^{24} \)), the variation of each \( N_k \) over consecutive integer values may be viewed as
an infinitesimal change in comparison to $N$. As such, we may consider the $N_k$ as continuous real variables, which then allows for differentiation.

Because the functional form of $W$ involves products and quotients, it is convenient to consider the logarithm of $W$, i.e.,

$$\ln W = \ln(N!) - \sum_{k=1}^{n} \ln(N_k). \tag{67}$$

Maximizing $\ln W$ is equivalent to maximizing $W$, since the logarithm function is strictly increasing, hence one-to-one, on $(0, \infty)$.

But there still remains the problem of how to deal with factorials. What is $\ln(N!)$? How can we work with it? The answer lies in a most important approximation of the logarithm of large numbers. It is known as “Stirling’s approximation:” For large $N$,

$$\ln(N!) \approx N \ln(N) - N. \tag{68}$$

This approximation gets better as $N$ increases.

Applying Stirling’s approximation to the $W$ function yields

$$\ln W = \ln(N!) - \sum_{k=1}^{n} [N_k \ln(N_k) - N_k], \tag{69}$$

where we have replaced the approximation sign with an equality. Note that the first term on the RHS of (69) is a constant, which will disappear after partial differentiation with respect to the $N_k$.

The Lagrangian associated with this optimization problem with two constraints will require two Lagrange multipliers, $\lambda$ and $\nu$. It is given by

$$L(N_1, N_2, \cdots, N_n, \lambda, \mu) = \ln(N!) - \sum_{k=1}^{n} [N_k \ln(N_k) - N_k] + \lambda(\sum_{k=1}^{n} N_k - N) + \mu(\sum_{k=1}^{n} N_k E_k - E). \tag{70}$$

We now look for critical points: For $i = 1, 2, \cdots, n$,

$$\frac{\partial L}{\partial N_i} = 0 \tag{71}$$

yields the following equations,

$$\ln(N_i) = \lambda + \mu E_i, \quad i = 1, 2, \cdots, n. \tag{72}$$

This implies that

$$N_i = e^{\lambda} e^{\mu E_i}. \tag{73}$$
If we now impose the first constraint (63),
\[ \sum_{i=1}^{n} N_i = e^{\lambda} \sum_{i=1}^{n} e^{\mu E_i} = N, \]  
we can eliminate the multiplier \( \lambda \),
\[ e^{\lambda} = \frac{N}{\sum_{i=1}^{n} e^{\mu E_i}}, \]
yielding the result,
\[ N_i = \frac{N}{\sum_{i=1}^{n} e^{\mu E_i}}. \]  

Normally, the next part of the exercise would be to determine the Lagrange multiplier \( \mu \). In some way, it could be connected to the total energy \( E \) appearing in the second constraint (64). The standard procedure, however, is to use the result for \( N_i \) in (76) in some expressions from Statistical Mechanics to make a connection with Thermodynamics. The result is that
\[ \mu = -\frac{1}{kT}, \]
where \( k \) is “Boltzmann’s constant” and \( T \) is absolute temperature. The final result for the \( N_i \) is then given by
\[ N_i = \frac{N}{\sum_{i=1}^{n} e^{E_i/kT}} e^{-E_i/kT}. \]  
This is often written in the form,
\[ p_i = \frac{N_i}{N} = \frac{e^{-E_i/kT}}{\sum_{i=1}^{n} e^{-E_i/kT}}, \]
where \( p_i \) denotes the fraction of atoms/molecules in state \( i \), so that
\[ \sum_{i=1}^{n} p_i = 1. \]
In whatever we choose to write it, these equations characterize what is known as the “Boltzmann distribution.”

In most applications, particularly those involving higher dimensions, e.g., \( \mathbb{R}^3 \), one must take into consideration the energy degeneracy of states, i.e., there are several ways that a state with a particular energy can be formed. For example, if we consider the kinetic energy of a molecule in \( \mathbb{R}^3 \), then there are many velocity vectors \( \mathbf{v} \) which have a given speed \( v \) – in fact, there is an entire sphere of such vectors, with radius \( v \).

As such, a weighting function must be introduced into the above formula. For simplicity, we simply write the form associated with the above discrete case:
\[ p_i = \frac{N_i}{N} = \frac{g_i e^{-E_i/kT}}{\sum_{i=1}^{n} g_i e^{-E_i/kT}}, \]
where \( g_i \) is the degeneracy of the \( i \)th state.

With some work, one can then apply this idea to determine the (continuous) distribution of speeds in an ideal gas. The complication is that we must now consider the components of the momenta/velocities of the gas molecules in three directions and how they combine to yield a velocity, therefore speed. We simply state the final result that the distribution of speeds, now considered as a continuous variable \( v \in [0, \infty) \) is given by the so-called “Maxwell-Boltzmann distribution” that you probably saw in your first-year Chemistry course. Some plots of this distribution for different temperatures \( T \) are shown below. (The plot was taken from the wikipedia.org site.)