Suppose \( f: \mathbb{R}^{n+1} \to \mathbb{R} \) and \( L \) is a curve or surface in \( \mathbb{R}^n \). Say we want to maximize (or minimize) \( f \) but only allowing points of \( L \) as inputs. I.e., we are wondering which points \( \mathbf{x} \in L \) give the largest value of \( f(\mathbf{x}) \).

Suppose \( \mathbf{x}_0 \in L \) is such a maximizing input. Then \( \nabla f(\mathbf{x}_0) \perp L \) at \( \mathbf{x}_0 \).

Theorem: \( \nabla f(\mathbf{x}_0) \perp L \) at \( \mathbf{x}_0 \)

(Note: This is not the theorem that says \( \nabla f(\mathbf{x}_0) \) is perpendicular to the level set of \( f \) at \( \mathbf{x}_0 \), because \( L \) is not a level set of \( f \) here! At most points \( \mathbf{x}_1 \in L, \nabla f(\mathbf{x}_1) \) will not be \( \perp \) to \( L \).)
Why is this true? If \( L \) makes an angle of \( \Theta < \pi/2 \) with \( \nabla f(x_0) \) at \( \bar{x}_0 \), then moving the input along \( L \) instantaneously in the direction of the tangent vector \( \bar{\mathbf{u}} \) (see picture) will change \( f \) at a rate of
\[
\nabla \bar{u} f(x_0) = \nabla f(x_0) \cdot \bar{u}
\]
\[
= |\nabla f(x_0)| \cdot |\bar{u}| \cdot \cos \Theta
\]
positive since \( \Theta < \pi/2 \).

This means \( f(x_0) \) can be increased slightly by moving \( \bar{x}_0 \) (inside \( L \)) in the (approximate) direction of \( \bar{u} \), so it is not yet maximized.

Now, if \( L \) is defined by implicit equations, often called "constraints," then we can find \( \bar{x}_0 \) by a method which depends on the number of constraints. For us, one or two.
1. If the set $L$ is defined by one implicit equation, say $g = c^2$, (so $L$ is a surface in $\mathbb{R}^3$ or a curve in $\mathbb{R}^2$), then we already know that $\nabla g(x_0)$ gives the normal direction to $L$ at $x_0$ (here we assume $\nabla g(x_0) \neq 0$), so $\nabla f(x_0) \parallel \nabla g(x_0)$, i.e.

$$\nabla f(x_0) = \lambda \nabla g(x_0)$$

Writing out the components of this equation, and remembering also that $g(x_0) = c$, we have $n$ equations in $n$ variables:

\[
\begin{align*}
g &= c \\
F_x &= \lambda g_x \\
F_y &= \lambda g_y \\
F_z &= \lambda g_z \\
\text{if in } \mathbb{R}^3 \wedge (F_z = \lambda g_z)
\end{align*}
\]

The scalar $\lambda$ is called a "Lagrange multiplier," and solving this system for $x_0, y_0, z_0$ is called "the method of Lagrange multipliers." The inputs $x_0$ you find ("Lagrange" inputs?) are like critical inputs: if you check them all, you will find the "true" $x_0 \in L$ which maximizes $F(x_0)$.

<-- Please read the text for examples of how to solve these equations... they can be quite tricky!
(2) If \( L \) is defined by two implicit equations, \( \{ g = c_1, \quad h = c_2 \} \), for example a curve in \( \mathbb{R}^3 \), then we know two normal vectors to \( L \) at \( \vec{x}_0 \) : \( \nabla g(\vec{x}_0) \) and \( \nabla h(\vec{x}_0) \). (here we assume \( \nabla g(\vec{x}_0) \) and \( \nabla h(\vec{x}_0) \) are not 0 and not parallel to each other). Since \( \nabla f, \nabla g, \) and \( \nabla h \) (at \( \vec{x}_0 \)) all lie in the plane of normal vectors to \( L \) at \( \vec{x}_0 \), we can express \( \nabla f \) as a "linear combination" of \( \nabla g \) and \( \nabla h \). there \( \phi \) \( \begin{align*} \nabla f &= \lambda \nabla g + \mu \nabla h \quad \text{at } \vec{x}_0 \\
abla g &= c_1 \\
abla h &= c_2 \\
x &= \lambda g_x + \mu h_x \\
y &= \lambda g_y + \mu h_y \\
z &= \lambda g_z + \mu h_z \end{align*} \) Now we have \( n+2 \) variables in \( n+2 \) unknowns, we can solve to find "Lagrange" inputs, \( \vec{x}_0 \), and check them to find the maximum.
Of course, if we can parametrize the curve \( L \) nicely, we can solve a one-variable calculus problem instead of a system of \( S \) equations, which is probably much easier. But parametrization is often extremely difficult or impossible in practice, so then Lagrange multipliers come to the rescue.
How two vectors "Span" a plane (optional)

Say \( \vec{u} \) and \( \vec{v} \) are non-zero, non-parallel vectors in \( \mathbb{R}^3 \) (or \( \mathbb{R}^n \)). Slide them into position at the origin, so they determine a plane:

Now imagine that this sheet of paper is that plane, and imagine a third vector \( \vec{w} \) in the same plane:

If we subtract (\( \pm \)) multiples \( \mu \vec{v} \) of the vector \( \vec{v} \) (\( \mu \) being a scalar), the tip of the resulting vector \( \vec{w} - \mu \vec{v} \) moves along the dotted line as \( \mu \) changes. For some \( \mu \), \( \vec{w} - \mu \vec{v} \) will line up with \( \vec{u} \), say \( \vec{w} - \mu \vec{v} = \lambda \vec{u} \) for some scalar \( \lambda \). Rearranging gives:

\[
\vec{w} = \lambda \vec{u} + \mu \vec{v}
\]