The method of Lagrange multipliers is used to find extrema of a function \( f \) restricted to level curves of some function \( g \). More precisely:

**Theorem 1.** Let \( g : \mathbb{R}^n \to \mathbb{R} \) be \( C^1 \) and let \( S = \{ x \in \mathbb{R}^n : g(x) = c \} \). Let \( f : \mathbb{R}^n \to \mathbb{R} \) be \( C^1 \) and suppose that \( f \big|_S \) (the restriction of \( f \) to \( S \)) attains a max or min at \( p \in S \). Then either \( \nabla g(p) = 0 \), or
\[
\nabla f(p) = \lambda \nabla g(p)
\]
for some scalar \( \lambda \).

**Proof.** Here is the proof in the case \( n = 2 \). The proof is similar for arbitrary \( n \).

Denote \( p = (x_0, y_0) \) and suppose \( \nabla g(x_0, y_0) \neq (0,0) \). Then either \( g_x(x_0, y_0) \neq 0 \) or \( g_y(x_0, y_0) \neq 0 \). Without loss of generality, suppose the latter. Then, since \( g(x_0, y_0) = c \), the Implicit Function Theorem implies that there is a neighborhood \( U \) of \( (x_0, y_0) \) and a \( C^1 \) function \( h \) defined on an open interval \( I \) containing \( x_0 \) such that \( h(x_0) = y_0 \) and
\[
S \cap U = \{ (x, h(x)), x \in I \}.
\]
Therefore
\[
g(x, h(x)) = c, \quad x \in I.
\]
Taking the derivative with respect to \( x \) gives
\[
g_x(x, h(x)) + g_y(x, h(x))h'(x) = 0
\]
Evaluating this at \( x_0 \) then implies
\[
g_x(x_0, y_0) + g_y(x_0, y_0)h'(x_0) = 0
\]
so
\[
h'(x_0) = -\frac{g_x(x_0, y_0)}{g_y(x_0, y_0)}.
\]
Now since \( f \big|_S \) attains its max/min at \( (x_0, y_0) \), the function \( f(x, h(x)) \) attains a max/min at \( x_0 \), so
\[
f_x(x_0, h(x_0)) + f_y(x_0, h(x_0))h'(x_0) = 0.
\]
Since \( h(x_0) = y_0 \), this implies
\[
f_x(x_0, y_0) - f_y(x_0, y_0)\frac{g_x(x_0, y_0)}{g_y(x_0, y_0)} = 0.
\]
Letting \( \lambda = \frac{f_x(x_0, y_0)}{g_y(x_0, y_0)} \), this implies
\[
f_x(x_0, y_0) = \lambda g_x(x_0, y_0) \quad \text{and} \quad f_y(x_0, y_0) = \lambda g_y(x_0, y_0),
\]
and thus \( \nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0) \).

\[\square\]
The possibility that $\nabla g(p) = 0$ cannot be ignored, as the following example demonstrates.

**Example 1.** Let $f(x, y) = x$, $g(x, y) = (x^2 + y^2 - 4)(x^2 + y^2 - 2x)$, and consider the level curve $S = \{(x, y) : g(x, y) = 0\}$. The level curve $S$ consists of the circles $x^2 + y^2 = 4$ and $(x - 1)^2 + y^2 = 1$. They are tangent at $(2, 0)$, and it is clear that this is the point in $S$ with the largest $x$-coordinate. Therefore $f$ attains its maximum along $S$ at $(2, 0)$. However, a little calculation reveals that

$$\nabla f(2, 0) = (1, 0) \quad \text{and} \quad \nabla g(2, 0) = (0, 0).$$

So at the maximum $\nabla g$ vanishes, but $\nabla f \neq \lambda \nabla g$ for any scalar $\lambda$.

There is a more general version of the Lagrange multiplier theorem which applies to the case of multiple constraints. We leave the proof to the reader.

**Theorem 2.** Let $g_i : \mathbb{R}^n \to \mathbb{R}$ be $C^1$ for $i = 1, \ldots, k$, $k < n$, and let $S = \{x \in \mathbb{R}^n : g_i(x) = c_i, i = 1, \ldots, k\}$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be $C^1$ and suppose $f|_S$ attains its max/min at $p \in S$. Then either $\nabla g_1(p), \nabla g_2(p), \ldots, \nabla g_k(p)$ are linearly dependent, or

$$\nabla f(p) = \lambda_1 \nabla g_1(p) + \lambda_2 \nabla g_2(p) + \cdots + \lambda_k \nabla g_k(p)$$

for some scalars $\lambda_1, \ldots, \lambda_k$. 

2