We want to study the maxima and minima of the function $f(x, y)$ with the constraint $g(x, y) = 0$. As you remember from one variable calculus, one usually finds the 'critical points' some of which may or may not be maximum or minimum, but all maxima and minima can be found among the critical points. (To decide maximum or minimum, one must use eithere the 'second derivative test' or some other means, so we will just try to find the critical points.) Also, rememeber, that calculus actually only gives 'local' maxima and minima.

As I said in class, one needs to parametrize the curve $g = 0$, at least near any given point in the region R of interest. So, for mathematical precision, one need to assume that at least one of $\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}$ to be non-zero at any point of R. This is the part which usually get swept under the rug in less rigorous discussions. (For example, if the region R contains the origin, and $g = x^2 - y^3$, then we will run into problems.) So, we will alwyas assume this. What is the advantage? This is known as implicit function theorem, a very important, useful and most-often quoted theorem.

Theorem 1. *Let $\frac{\partial g}{\partial y} \neq 0$ at a point $P = (x_0, y_0)$ where $g(P) = 0$. Then in a neighbourhood of P, we can solve y as a function of x. That is there exists a function $y = \phi(x)$ near P, with $y_0 = \phi(x_0)$, so that $g(x, \phi(x)) = 0$ for all x in an interval containing x_0.*

Proof. This is just a sketch. By multiplying g by a non-zero constant, we may assume $\frac{\partial g}{\partial y}(P) = 1$. Then we solve for ϕ by approximation. For simplicity of notation, let me also assume that P is the origin. What happens if we put $y = -\frac{\partial g}{\partial x}(P)x + x^2\phi_1(x)$ where ϕ_1 is to be determined? One can easily check using the Maclaurin expansion, that $g(x, y(x))$ looks like $x^2 g_1(x)$ for some g_1. Using g_1, one can solve for a suitable ϕ_1 so that now if we put $y = -\frac{\partial g}{\partial x}(P)x + ax^2 + \phi_2(x)x^3$, then $g(x, y(x)) = x^3 g_2$. We continue this way to get a series expansion for y as a function of x. (Look up a book on Advanced Calculus or Analysis for more details. Of course, we are tacitly assuming that g has all derivatives here, but there are simpler proofs where only the first derivatives are used.)

$\square$

The more general version is as follows, which can be used to show Lagrage multiplier results for more constraints in more variables.

Theorem 2. *Let $g_1(x_1, \ldots, x_n), g_2(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n)$ be m functions with $m < n$ and the property that if $P = (a_1, \ldots, a_n)$ is a point in n-space where all the g_i's are zero, then the Jacobian matrix*

$J(g_1, \ldots, g_m)$ defined as the matrix whose ijth entry is $\frac{\partial g_i}{\partial x_j}$ has rank m (the maximum possible) at P. So, we assume without loss of generality that the first $m \times m$ minor of this matrix does not vanish at P. Then there exists functions $\phi_i(x_{m+1}, \ldots, x_n)$ for $1 \le i \le m$ so that, $a_i = \phi_i(a_{m+1}, \ldots, a_n)$ and $g(\phi_1, \phi_2, \ldots, \phi_m, x_{m+1}, \ldots, x_n) = 0$ in a neighbourhood of P.

As an example, let us take $m = 1$ and let $g = x_1^2 + x_2^2 + \cdots + x_n^2 - 1$ and $P = (1, 0, \ldots, 0)$. Then, $\frac{\partial g}{\partial x_1} \neq 0$ at P (and all others are zero). Then it is clear that we can take $\phi_1 = \sqrt{1 - x_2^2 - \cdots - x_n^2}$, where we mean the positive square root, near P. You can see that if our point is $P = (-1, 0, \ldots, 0)$, then we should have taken the negative square root function. You can see that at different points, we may need different functions and only certain variable can be written in terms of the others.

Theorem 3. *The following are equivalent, with notation as above.*

1. *A point $P = (x_0, y_0)$ is a critical point of f with constraint $g = 0$.*
2. *$g(P) = 0$ and $\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x} = 0$ at P.*
3. *There exists a λ so that $g(P) = 0$, $\frac{\partial f}{\partial x}(P) + \lambda\frac{\partial g}{\partial x}(P) = 0$ and $\frac{\partial f}{\partial y}(P) + \lambda\frac{\partial g}{\partial y}(P) = 0$.*

Proof. Let us show that the first implies the second. By assumption, we have at least one of $\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}$ is not zero at P. so let us assume that $\frac{\partial g}{\partial y}(P) \neq 0$, the other case being similar. Then we have a $\phi(x)$, so that $\phi(x_0) = y_0$ and $g(x, \phi(x)) = 0$ at all points near x_0. So, the curve $g = 0$ is defined as $y = \phi(x)$ near this point. Thus, we are looking at the one variable function $f(x, \phi(x))$, whose critical points we seek. But, P is a critical point implies that $\frac{df(x,\phi(x))}{dx} = 0$ at x_0. Expanding, we get,

$$\frac{\partial f}{\partial x}(x, \phi(x)) + \frac{\partial f}{\partial y}(x, \phi(x))\phi'(x) = 0$$

at $x = x_0$. We also have,

$$\frac{\partial g}{\partial x}(x, \phi(x)) + \frac{\partial g}{\partial y}(x, \phi(x))\phi'(x) = 0$$

at x_0. Solving for ϕ' from the second and substituting in the first, we get the second statement.

To go from the second to the third, if we take $\lambda = -\frac{\partial f}{\partial y}(P)/\frac{\partial g}{\partial y}(P)$, (this is allowed, since $\frac{\partial g}{\partial y}(P) \neq 0$), then it is easy to see that the three equations are statisfied once you have (2).

Finally we will show that (3) implies (2), since we have not precisely defined (1). As before we substitute for λ from the last equation in the last but one to get, $\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x} = 0$ at P. $\square$

Here is the more general version.

Theorem 4. *Let f be a function in n variables and $0 = g_i, 1 \leq i \leq m < n$ be the constraints satsfying the property stated in the implicit function theorem. Then a point $P = (a_1, \ldots a_n)$ is a critical point of f with the given constraints if and only if there exists $\lambda_1, \ldots, \lambda_m$ such that these satisfy the $n + m$ equations, $\frac{\partial f}{\partial x_i} + \sum \lambda_k \frac{\partial g_k}{\partial x_i} = 0, 1 \leq i \leq n$ and $g_k = 0, 1 \leq k \leq m$ at P.*