Several students have asked me how one can classify the critical points that are found using Lagrange Multipliers, that is, how does one know that a critical point is a (local) minimum, or maximum, or saddle point, or something else. That's an excellent question; I'm glad you thought to ask. But it does not have an easy answer. There are several ways you can think about it. I'd like to show you some of them — they will encourage you to tie together a lot of separate ideas we have already developed!

Allow me to illustrate with one of the problems that appear in the homework:

$$\text{Minimize } \; f(x,y,z) = 4x^2 + 3y^2 + 2z^2 \text{ subject to } \; 2x + 3y + z = 9$$

The Lagrange Multipliers Theorem tells you that if f attains a (local) minimum or maximum at the point P, then the gradient of f at the point P will be parallel to the gradient of $g(x,y,z) = 2x + 3y + z - 9$ at the point P. That means the coordinates of any such point will satisfy

$$\begin{aligned} 8x &= 2\lambda \\ 6y &= 3\lambda \\ 4z &= \lambda \\ 2x + 3y + z &= 9 \end{aligned}$$

for some real number λ. That's a system of 4 (linear!) equations in 4 unknowns, which I know you can solve; the only solution is the point

$$x = 1, \quad y = 2, \quad z = 1, \quad (\text{and } \lambda = 4)$$

But what happens at this point? Does f attain a minimum there? The short answer is, yes it does, but there's no single method that will easily prove this to you in this and every similar case. You sometimes have to be creative in your analysis.

1 For example, here's a completely algebraic way to answer this question. Consider the function

$$h(x,y,z) = 4(x-1)^2 + 3(y-2)^2 + 2(z-1)^2 + 18$$

I will let you expand this if you want; it's the sum of seven terms, three of which sum to the original function f and the other four sum to exactly $8x + 12y + 4z - 36 = 4g$, that is,

$$f(x,y,z) = h(x,y,z) + 4g(x,y,z)$$

The significance of this equation is that *as long as you stay on the set where* $g = 0$, this function h will take exactly the same values as the original function f. On the other hand, it is clear from the presentation of h as a sum of squares that it will always take on values that are larger than 18, except right at our critical point, where $h(P) = f(P) = 18$. This proves that f achieves its minimum value at P, among all points where $g = 0$.

That was a little bit of magic, though: it required that I be able to add a multiple of g to f so as to get a function which is positive on all of Euclidean space. I'm not sure when that's possible, and I don't have an algorithm that would help you find such a multiplier when it *is* possible, although it is not an accident that the multiplier of 4 in the previous paragraph is the same as the value of λ at the critical point. (It is, sort of, true that you will recognize an always-positive function in the same way I just did, namely, recognizing it as a sum of squares. This is a subtle bit of research mathematics; look up *Hilbert's Seventeenth Problem*.)

2 A second idea is also computational, and it builds on the following idea. The condition that there be a critical point at P in this constrained problem may be described in this way: P is a point that makes all the partial derivatives of $F = f - \lambda g$ vanish — *including the partial derivative with respect to* λ. (Observe that F is a function of *four* variables, namely x, y, z, λ, so setting *all* the partials to zero gives you the four equations in four unknowns that you need.) If you take this perspective, then the condition to have a minimum is easy to describe: it is that the second derivative ("Hessian") matrix must indicate that F has a minimum at this point in $\mathbf{R}^4$.

In our example, $F = 4x^2 + 3y^2 + 2z^2 - \lambda(2x + 3y + z - 9)$ has gradient

$$\langle 8x - 2\lambda,\ 6y - 3\lambda,\ 2z - \lambda,\ -(2x + 3y + z)\rangle$$

and hence the Hessian matrix is

$$\begin{pmatrix} 8 & 0 & 0 & -2 \\ 0 & 6 & 0 & -3 \\ 0 & 0 & 2 & -1 \\ -2 & -3 & -1 & 0 \end{pmatrix}$$

We have hardly mentioned second derivatives except for functions $\mathbf{R}^2 \to \mathbf{R}$ but you can see they will also be symmetric matrices. In order to test that they reveal a local minimum, one must check all the smaller square matrices that include the upper left corner:

$$(8), \quad \begin{pmatrix} 8 & 0 \\ 0 & 6 \end{pmatrix}, \quad \begin{pmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 8 & 0 & 0 & -2 \\ 0 & 6 & 0 & -3 \\ 0 & 0 & 2 & -1 \\ -2 & -3 & -1 & 0 \end{pmatrix}$$

The theorem about unconstrained optimization is that if all these matrices have positive determinants, then the function has a local minimum at the critical point; if they alternate in sign starting with a negative, then there is a local maximum. Any other mix of signs indicates some type of saddle point, and any determinant being zero suggests there can be some kind of degeneracy.

Well, exactly the same kind of conclusion may be used in the case of a constrained optimization problem, and in this example it proves that our original function f attains a local minimum at P (compared to all the other points where $g = 0$). That's useful, although it doesn't actually prove that f attains an *absolute* minimum at P. Even in the unconstrained case, the fact that a local minimum is the only critical point does not guarantee that the function attains an absolute minimum there. An example is the function

$$f(x, y) = x^2 + y^2(1 - x)^3$$

which attains a local minimum value of 0 at the origin and has no other critical points at all, and yet $f(4,0) = -11 < f(0,0)$ so the origin is clearly not a global minimum.

3 Another approach you could use to address constrained optimization problems is to make them unconstrained, in the following sense. In our example, we are supposed to consider only those points (x, y, z) which make $g(x, y, z) = 0$. This surface is easily seen to be a plane, described implicitly. You can describe it explicitly too of course: it's the graph of

$$z = 9 - 2x - 3y$$

In other words, we use the equation $g = 0$ to solve for one variable in terms of the others. Then we can substitute this into the description of f: let

$$f_2(x, y) = 4x^2 + 3y^2 + 2(9 - 2x - 3y)^2$$

Then this f_2 is defined on all of $\mathbf{R}^2$, and we need only to minimize it; find the point $p \in \mathbf{R}^2$ where f_2 is minimized, then compute a z coordinate from the equation $g = 0$ as above, to find the point $P \in \mathbf{R}^3$ where f will be minimized on the surface $g = 0$. Of course to find p we simply compute the gradient of f_2 to get $\langle 8x - 8(9 - 2x - 3y), 6y - 12(9 - 2x - 3y)\rangle$ which only vanishes at $p = (1, 2)$; then $z = 9 - 2x - 3y = 1$ so $P = (1, 2, 1)$ as before. But in this perspective the problem is easier: we now know we can test the behaviour of this critical point by computing the Hessian of f_2. I get $\begin{pmatrix} 24 & 24 \\ 24 & 42 \end{pmatrix}$, which indicates f_2 has a local minimum at p, and hence f has a local minimum (among points with $g = 0$) at P.

Notice that this approach requires viewing the surface $g = 0$ as the graph of a function $z = 9 - 2x - 3y$. You don't really have to explicitly solve for z, you just need to know that it can be done and you need to be able to compute $\partial z/\partial x$ and $\partial z/\partial y$ in order to locate and then classify the critical points of f_2. Guess what: this is why I gave you the Implicit Function Theorem! It says that we can indeed do this near any point where $\partial g/\partial z$ is not zero (in our example, that's everywhere), and it tells us that the gradient of z is exactly

$$\frac{\partial z}{\partial(x, y)} = -\left(\frac{\partial g}{\partial z}\right)^{-1} \left(\frac{\partial g}{\partial(x, y)}\right)$$

But perhaps that observation is best saved for another day.

4 Another option you might prefer, rather than trying to invent a second-derivative test for constrained optimization, is to remember that in 1-variable calculus we also had a *first*-derivative test to classify critical points. Clearly if a function increases to the left of $x = a$ and decreases to the right of $x = a$, then it has a local max at $x = a$. In the same way, we could simpliy ask: does our function decrease as we move towards P from any nearby point Q that lies on the plane $g = 0$? For that task we can use directional derivatives: the rate at which f changes (instantaneously) as we from Q in the direction $v = P - Q$ is $D_v f(Q) = \nabla f(Q) \cdot v$. In this particular example we can easily calculate ∇f to be the vector-valued function $\nabla f(x, y, z) = \langle 8x, 6y, 4z\rangle$ and then evaluate it at this nearby point $Q = P - v$. If we give names to the components of v, say $v = \langle v_1, v_2, v_3\rangle$, then $\nabla f(Q) = \langle 8(1-v_1), 6(2-v_2), 4(1-v_3)\rangle$ and so we compute the directional derivative as we head from Q to P with a dot product:

$$D_v f(Q) = \nabla f(Q) \cdot v = (8v_1 + 12v_2 + 4v_3) - (8v_1^2 + 6v_2^2 + 4v_3^2)$$

BUT: remember that we are only comparing the values of f at points in the plane where $g = 0$, so v will be a vector in that plane, i.e. it will be perpendicular to the normal vector $\langle 2, 3, 1\rangle$ of the plane, which is exactly what we need to make the linear part of this directional derivative vanish. Hence for such vectors v we will have $D_v f(Q) = -(8v_1^2 + 6v_2^2 + 4v_3^2)$, which is clearly negative: so f will decrease as we move toward P within the plane.

5 If you try to work out the pattern of the preceding paragraph for general functions f and general constraints g, you soon get the idea to just use the initial terms of the Taylor series of f near P. You know

$$\nabla f(P + v) \approx f(P) + \nabla f(P) \cdot v + \frac{1}{2} v^t f''(P) v$$

Apply this to vectors v that point along the surface $g = 0$, to see how f changes if you move to other points in that surface. Doing so restricts you to vectors that are perpendicular to $\nabla g(P)$. If P is a point selected by the Lagrange Multipliers condition, then this will mean v is also perpendicular to $\nabla f(P)$, so the linear term in this Taylor expansion is exactly 0. So the analysis comes down to deciding whether the expression $v^t f''(P) v$ is positive or negative for such vectors v. In our particular case, this expression works out to $8v_1^2 + 6v_2^2 + 4v_3^2$, which is obviously positive for every vector v, whether it points along the surface or not; thus f will increase if you move in any direction along the plane $g = 0$. In a more general setting this might require a little more Linear Algebra (we are deciding whether a quadratic form is positive-definite on a particular subspace) but I can give you some pointers if you are interested.

6 One more idea is more geometric. If you use the Lagrange Multiplier Theorem, you can isolate all candidates for where a function might attain a minimum or a maximum. Separately, you might be able to use a different theorem that assures you that there IS a minumum somewhere (and a maximum somewhere else). I gave you such a theorem: there will always be a min and a max as long as f is continuous and the domain of options is both "closed" and "bounded". In our constrained optimzation problems, the domain is always the set of points where $g = 0$; as long as g is also continuous, this set will be closed (i.e. it will include its boundaries). For example, our domain is that plane in $\mathbf{R}^3$; it doesn't even *have* any boundaries to include, so it's a closed set by default.

Unfortunately the plane is not bounded, so we cannot use that theorem as it stands. Nonetheless we can adapt it as follows. Recall that our f attained a fairly low value of 18 at our critical point P; I claim that the function f never attains a value lower than 18 once we are outside the ball of radius 3 centered at the origin. That's kind of obvious: if you're outside that ball, then $x^2 + y^2 + z^2 > 9$, and so $f = 4x^2 + 3y^2 + 2z^2$ will be at least as big as $2x^2 + 2y^2 + 2z^2$, which in turn will have to be larger than 18, and hence not lower than the value of f at our P. So really it is sufficient to look at the points in the plane which happen to lie inside the closed ball of radius 3. Those points will lie in a disk (any plane meets any ball in a disk if it meets the ball at all), so now we are trying to minimize f on a set which is closed *and bounded too* and hence from that theorem we know f will attain a minimum somewhere in there. Lagrange Multipliers shows us there is only one candidate point, so that must be where f attains an absolute minimum.

(Use this line of reasoning carefully. The same theorem shows that f must attain a *maximum* somewhere on the disk; the Lagrange Multiplier theorem says the maximum must occur at a place where ∇f and ∇g are parallel, OR at a point where one of the gradients does not exist, OR at point on the boundary of this disk. The only place where the gradients line up is our critical point P where the minimum occurs; and there are no points without a gradient. But we DO have to look at points on the circle that forms the boundary of the disk, if we want to find the maximum of f. I won't do it because although that point gives the max of f on the disk, it does not give the max of f on the whole plane: the inequalities I used in the last paragraph work against us when searching for a max instead of a min.)

I'm sure there are other ways to address the problem of characterizing what exactly happens at a critical point in a constrained optimization problem, but I decided to highlight these in order to (a) give you some tools you might want to reach for on a future problem, and (b) to illustrate how everything else we've done up to this point can be used to our advantage to make sense of what's going on in one example!