MA 261 **On LAGRANGE MULTIPLIERS** *based on Stewart's text*

We assume that you are familiar with the standard tests for extrema when we have a function f of more than one independent variable; we usually study $f(x, y)$ but the same ideas work for $f(x, y, z)$, etc. The important thing is being clear on the difference between **independent** *and non-independent variables.* In this introduction we and examples we write the variable as (x, y) or (x, y, z), but use vector notation in describing the general principle.

1 Independent Variables

Suppose we want to find the extrema of $f(x, y)$ where (x, y) ranges over some domain D (for convenience we always assume that f has continuous partials). The standard method is (a) to identify the possible interior extrema by finding points (x, y) where $\nabla f(x, y) = \mathbf{0}$. This is where the issue of the variables being independent arises, since when they are independent, we may leave (x, y) in any direction and check whether f increases or decreases in that direction (that is why directional derivatives are introduced.) we introduced directional derivatives in the previous lecture). Only if $\nabla f(x, y) = \mathbf{0}$ will the *directional derivative* of f in any direction from (x, y) be zero, which means that there is no obvious direction in which to go from (x, y) where f would increase or decrease, and so (x, y) has to be considered a point at which f might attain an extremum.

Step (a) is followed by (b): identifying the possible extrema taken on the boundary of D, the situation discussed here.

2 Dependent variables

The situation is different when we look for points on the boundary of D at which the extrema occur; this is analogous to checking at the 'endpoints' when f is defined on $[a, b]$, and f might have an extremum at either $x = a, x = b$ without $f'(a)$ or $f'(b)$ vanishing. In two variables this usually that means that a given portion of ∂D (the boundary of D) is represented by one of the forms (i) $y = g(x)$, (ii) $x = g(y)$, or (iii) $(x, y) = (x(t), y(t))$, so that t is a parameter.

It is important to see in that situation we have lost an independent variable! In case (i) we are saying that only x is independent, so that once x is given we know y – that can't occur were the variables inependent. In (ii) y is the only independent variable, and in (iii) only t is independent. So, for example, in (i), we have $f(x, y) = f(x, g(x)) = F(x)$, in case (ii) $f(x, y) = F(y)$, and in (iii) $f(x, y) = f(x(t), y(t)) = F(t)$ [I am using the letter F for a different function in each case]. When $(x, y) \in \partial D$, the variables are not independent, so we are not able to move (x, y) in any direction we choose and remain on the boundary. That is the reason f might have an extremum at (x, y) without $\nabla f(x, y)$ being zero.

(When we consider a function f of n variables defined in a region D, there will in general be $n-1$ independent variables on ∂D.)

In our earlier homework, when doing these problems without Lagrange multipliers, we would simply study f on this portion of ∂D by going back to one-variable maximum/minimum techniques from calculus, and so identify all possible extrema on the boundary. This can lead to many cases, when the boundary has pieces which are of various types (i)–(iii). [There will be examples given.]

3 Naïve way to handle boundary situation

The Lagrange method is a more insightful – and simple – way of handling the situation that the number of independent variables is reduced — that is what we mean by there being a *constraint*. As we have mentioned, the constraint is that either $y = g(x), x = h(y)$ or $x = x(t), y = y(t)$, etc., which as we have observed reduces the number of independent variables.

4 Lagrange's insight

Lagrange had an insight that uses geometry and some elementary vector analysis. The exposition I am giving here covers §14.8 in (my opinion) a simpler way, and always leads to "one less equation" in a system of simultaneous equations. No matter how we approach these problems, solving these systems requires methods improvised for each problem – we do not use any general method.

Our problem is to

$$\text{extremize } f(\mathbf{x}) \text{ subject to the `constraint' } g(\mathbf{x}) = 0:$$

here $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ is a vector.

It is governed by an elementary fact about vectors.

Principle: *Let $\mathbf{x} = (x_1, x_2 \ldots, x_k)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_k)$ be be vectors (in our context usually $k = 2$ or 3.) Then $\mathbf{x} \parallel \mathbf{y}$ if and only if all ratios $x_i/y_i (i = 1, 2, \ldots, k)$ are the same; the understanding is that if we have $x_j = 0$ for some j, then the corresponding entry y_j must also be zero.*

(You should check with a few examples: for what λ is $(1, 4, 9) \parallel (-2, -8, \lambda)$? How would your answer change if you replace only 4 with 6? Or 9 with 0?)

Let us apply this to the situation of extremizing f subject to our constraint $g = 0$, and let $\mathbf{x} = (a, b)$ be a point we are testing as a possible extrema. Then **Lagrange:** in order that f have an extrema at $\mathbf{x}$ it is necessary that

$$\nabla f(\mathbf{x}) \parallel \nabla g(\mathbf{x}). \tag{1}$$

NOTE. The word *multiplier* arises from the usual formulation of this principle: it asserts that at a potential extrema we have the equation

$$\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}),$$

where λ is a scalar. You should check that these two ways of expressing the Lagrange principle are the same; the (slight) advantage in the formulation I prefer is that the scalar λ usually has no physical significance, and in practice (that is, homework problems!) means introducing λ simply gives is an extra equation to consider, with no extra information.

Once we identify the points where (1) holds, we have to compute f at each, and then the maximum of f among points which satisfy the constraint will be where f takes on the largest value among these points; a similar remark will lead to the minimum of f.

5 Some examples

Let's see how this works with some of the homework and examples in the text. We should always identify f and g. We then use the Lagrange principle, and what will come out is that the coordinates at a possible extremum must satisfy a certain *relation*. In other to find the exact values of the coordinates at these possible extrema, however, we have to return to the constraint $g = 0$, since if our constraint were $g = c$, c a constant (with $c \neq 0$), the Lagrange equation would be the same, and so would be a *relation* between the variables. (Our first example is an exception to this principle, because the relation (1) is more subtle, since the coordinates at the extrema will be zero.) I start with a complicated example, which was done earlier by other methods.

P. 962, No. 39. Find a point on the surface $z^2 = xy + 1$ closest to the origin.

Step I**:** Find f, g. Well, it is f that we wish to minimize, so

$$f = (x-0)^2 + (y-0)^2 + (z-0)^2 = x^2 + y^2 + z^2;$$

where x, y, z are restricted to the surface; thus

$$g(x, y, z) = z^2 - xy$$

(the constraint is $g = 1$).

Step II**:** Write the Lagrange equation for this case; we can cancel common factors. This leads to

$$\nabla f \parallel \nabla g : (x, y, z) \parallel (-y, -x, 2z). \tag{2}$$

(Note that without the constraint of $g = 1$, there is a trivial solution: f has an absolute minimum when $x = y = 0$, but the points $(0, 0, 0)$ does not satisfy the constraint $z^2 = 1 + xy$.)

Step III**:** Using our Principle, manipulate (2) to get a *relation* that must be satisfied at a possible extremum.

I always assume first that $(*)$ none of the coordinates are zero. So if these vectors are parallel, then (looking at the first two coordinates only) we see that $x/y = y/x$ so $x = \pm y$ which means that $x/y = \pm 1$. But then that common ratio

(± 1) of the first two coordinates in (2) would have to be the same for the ratio of third coordinates, and so the same as $2z/z$; however, unless $z = 0$, $2z/z$ can't be ± 1. In other words, we are at a dead end!

This means the extrema will occur when $z = 0$ or, when we recall the case excluded in ($*$), that $x = 0$; if $x = 0$, then by our principle concerning parallel vectors, we'd have $y = 0$ too.

This means the only possible points to check are $(0, 0, \pm 1)$ (the ± 1 is the value of z when $x = y = 0$), and in addition, returning to the case that $z = 0$ we would have as possibilities the points $(1, -1, 0)$ and $(-1, 1, 0)$. At these points we compute $f : x^2 + y^2 + z^2$, and see that the extrema occur at $(0, 0, \pm 1)$: the closest point(s) on the surface are distance 1 from the origin.

P. 966, Example 1. Here $f = xyz$, and $g = 2xz + 2yz + xy$ (so the constraint is that $g = 12$).

Step *II***:** We see that $\nabla f \parallel \nabla g$ when

$$(yz, xz, xy) \parallel (2z + y, 2y + x, 2x + 2y).$$

When we divide, it is simplest to put the entries of the left-hand vector in the denominators, since the algebra is simpler. Indeed, if x, y or z is zero, the box has no volume, so we may assume that $xyz \neq 0$.

Step *III***:** Looking at the (ratio of) first two coordinates and cross-multiplying, we find that $2xz^2 + xyz = 2y^2z + xyz$, or that $x = y$. Then if we look at the (ratio of) first and third coordinates in the same way, we find that $2xyz + xy^2 = 2xyz + 2y^2z$, or, more simply, $x = 2z$. In summary, $x = y = 2z$, and since $2xz + 2yz + xy = 12$, we find that $x = 2 = y, z = 1$. (Notice that you get a different answer if $2xz + 2yz + xy = 100$ we would have a box of different size, but as we mentioned at the beginning of the §, the same relations between the variables x, y, z would remain.

P. 968, Example 3. Now $f = x^2 + 2y^2, g = x^2 + y^2 - 1$ (when we look inside the circle of radius one, x and y are independent variables, and the Lagrange method is not relevant there).

Step *II***:** $\nabla f \parallel \nabla g$ means that we have

$$(x, 2y) \parallel (x, y).$$

So now we see that the extrema occue when either x or y is zero, otherwise we'd have nonsense thinking the vectors parallel: $2 \neq 1$. So we need consider f only at $(0 \pm 1), (\pm 1, 0)$.

P. 971, No. 9 (This is a problem in three variables:) Maxtimize $f(x, y, z) = xyz; g : x^2 + 2y^2 + 3z^2 = 6$) So $\nabla f \parallel \nabla g$ occurs when

$$(yz, xz, xy) \parallel (x, 2y, 3z).$$

If $xyz = 0$, then $f = 0$. So if $xyz \neq 0$, we may put the entries of the second vector in the denominators, and see that an extemum will occur when

$$x^2z = 2y^2z;\ x^2y = 3z^2y.$$

That gives the relation that $x = \pm y, x = \pm\sqrt{3}z$. We now use that $g = 6$; note again that if the constraint were that $g = 100$, we would have the same relation between x, y, and z, but the specific number would be different. This produces eight points where extrema may occur, dependingon the choice of sign for each of the coordinates. A minimum will occur when exactly one or three of x, y, z is negative and the other positive, and a maximum occurs when either all three variables are positive or exactly two are negative.