Lecture 4-1

We continue with the method of Lagrange multipliers for maximizing or minimizing functions of several variables defined only on (or restricted to) level sets of functions. What we are doing, in the words of the title of §16.7, is computing maxima and minima with side conditions; this is sometimes also called computing constrained maxima and minima.

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and the level set S corresponding to the function $g : \mathbb{R}^n \to \mathbb{R}$ and the constant c, we know that the equation $g(x_1, \dots, x_n) = c$ can be solved differentiably for any of the variables x_i in terms of the others near a point $\vec{a} = (a_1, \dots, a_n)$ such that $(a_1, \dots, a_n \in S$ (so that $g(a_1, \dots, a_n) = c$) provided that $\partial g/\partial x_i(\vec{a}) \neq 0$; if we do this and insert the resulting expression for x_i into the formula for f, we obtain a function of the x_j for $j \neq i$ (still denoted f, following our usual sloppiness in notation). If this function has a local maximum or minimum at $\vec{a}$, its x_j partial, given by $\partial f/\partial x_j(\vec{a}) + (\partial f/\partial x_i)(\vec{a})(\partial x_i \partial x_j)(\vec{a}) = \partial f/\partial x_j(\vec{a}) - (\partial f/\partial x_i(\vec{a}))\frac{-\partial g/\partial x_j(\vec{a})}{\partial g/\partial x_i(\vec{a})} = 0$, whence $(\partial f \partial x_j)(\partial g/\partial x_i)(\vec{a}) - (\partial f/\partial x_i)(\partial g/\partial x_j)(\vec{a}) = 0$. The only way that this can hold for all $j \neq i$ and for all i with $\partial g/\partial x_i(\vec{a}) \neq 0$ is to have either $\nabla g(\vec{a}) = \vec{0}$ or $\nabla f(\vec{a}) = \lambda \nabla g(\vec{a})$ for some $\lambda \in \mathbb{R}$. This then is the criterion for $\vec{a}$ to be a critical point of f when this function is restricted to S. The scalar λ is called a *Lagrange multiplier*. Note that the two possibilities $\nabla g(\vec{a}) = \vec{0}$ and $\nabla f(\vec{a}) = \lambda \nabla g(\vec{a})$ can be captured by the single condition that $\nabla f(\vec{a}) \times \nabla g(\vec{a}) = \vec{0}$, if $n = 3$. Note also that ultimately we (usually) do not care about the value of λ; it is just a tool that we use to solve max-min problems.

In practice, just as we have neglected the possibility that $\nabla f(\vec{a})$ is undefined in computing critical points of a function f, so we also neglect the possibility that $\nabla g(\vec{a}) = \vec{0}$ as a critical point $\vec{a}$ in this context, concentrating our attention (at least in the examples we consider) on smooth points of the level set S. What we will find in most cases is that constrained maxima and minima occur at points that are highly symmetric in some sense. For example, returning to the problem discussed in an earlier lecture of maximizing $f(x, y) = xy$ subject to the constraint $g(x, y) = x^2 + y^2 = 1$, we find that the Lagrange multiplier condition $\nabla g = \lambda \nabla f$ translates to $2x = \lambda y, 2y = \lambda x$ for some λ, whence $y^2 = x^2 = 2\lambda xy$, forcing either $x = y = \pm\sqrt{2}/2$ or $x = -y = \pm\sqrt{2}/2$. (Notice that we had to continue to bear the constraint $g(x, y) = 1$ in mind even after we had set up the Lagrange equations.) We will not attempt to determine the nature of all of these critical points (this requires techniques beyond the scope of this course) but we will compare the values of f at all critical points to find the absolute maximum and minimum. Doing this, we find that the maximum value of $1/2$ occurs at $\pm(\sqrt{2}/2, \sqrt{2}/2)$ while the minimum value of $-1/2$ occurs at $\pm(\sqrt{2}/2, -\sqrt{2}/2)$, in accordance with our earlier calculation, which exploited the parametrization $(\cos t, \sin t)$ of the level set.

Next we look at the very classical problem of maximizing the product $x_1 \dots x_n$ of n variables x_i subject to the constraints that $x_i \geq 0$ for all i and $\sum x_i = 1$. Note before we compute any partial derivatives that this product can be made arbitrarily small (though positive) by letting one of the variables approach 0; so we know in advance that if there is just one critical point it must be the unique maximum. Next, Lagrange multipliers tell us that we must have for every index i that the product of the x_j for $j \neq i$ equals λ for some fixed λ, whence on multiplication by x_i we deduce that all x_i must be equal; their common value by the constraint is $1/n$. More generally, *the arithmetic mean* $(1/n)\sum_{i=1}^n x_i$ *of* n *positive numbers is always greater than or equal to their geometric mean* $(\prod_{i=1}^n x_i)^{1/n}$, *with equality if and only if all the numbers are equal*; this the famous *arithmetic-geometric mean inequality*, which is a workhorse in solving many problems involving inequalities.

Turning now to a problem whose solution is less symmetric than the previous ones, we minimize the surface area $xy + 2xz + 2yz$ of a box of dimensions x, y, z without a top of specified volume 12 cubic feet (Example 5 in the text, p. 846). Here the Lagrange equations become $y+2z = \lambda yz, x+2z = \lambda xz, 2x+2y = \lambda xy$. Multiplying the first equation by x, the second by $-y$, and adding, we get $2z(x-y) = 0$, which since $z \neq 0$ implies that $x = y$. Replacing y by x in the third equation we get $4x = \lambda x^2$, which since $x \neq 0$ implies that $x = y = 4/\lambda$; then the first equation shows that $z = 2\lambda$. Finally, plugging in the condition $xyz = 12$ we get $\lambda = \frac{2}{\sqrt[3]{3}}, x = y = 2\sqrt[3]{3}, z = \sqrt[3]{3}$. This unique critical point must be a minimum as it is easy to see that the surface area can be made arbitrarily large by making z very small. Hence the box with least surface area has the same length as width, but height half the length or width.

Finally, we let $A = (a_{ij})$ be a *symmetric* $n \times n$ matrix, so that $a_{ij} = a_{ji}$ for all indices i, j. We consider the problem of maximizing or minimizing the quantity $f(\vec{x}) = (\vec{x})^t A\vec{x}$ with constraint $||\vec{x}|| = 1$, where $\vec{x} = (x_1, \dots, x_n)$ is a *column* vector of variables x_i and $\vec{x}^t$ is the same vector written as a row vector. Then we have $f(\vec{x}) = \sum_{i=1}^n a_{ii}x_i^2 + 2\sum_{i,j=1,i<j}^n a_{ij}x_i x_j$ by the definition of matrix multiplication; our constraint is equivalent to requiring that $\sum x_i^2 = 1$. The Lagrange multipliers tell us that the condition to be a critical point is that $\sum_{j=1}^n 2a_{ij}x_j = 2\lambda x_i$, for all indices i. This last condition says exactly that $\vec{x}$ is a so-called *eigenvector* for A, meaning that $A\vec{x} = \lambda\vec{x}$; we call λ the corresponding *eigenvalue*. Here the existence and value of λ are much more interesting than the solution to the problem. We will explore this situation in more detail later.