Math 1c	TA: Padraic Bartlett
Recitation 4: Lagrange Multipliers and Constrained Optimization	
Week 4	*Caltech 2012*

1 Random Question

(This lecture ran a bit too long to get to the random question. It will reappear next week!)

2 Lagrange Multipliers

2.1 Statement of the method.

In last week's recitation, we talked about how to use derivatives to find and classify the **critical points** of functions $\mathbb{R}^n \to \mathbb{R}$, which we did using the following blueprint:

- First, look at $D(f)$, the matrix of derivatives of our function f. The critical points of f are precisely the points $\mathbf{a}$ at which $D(f)\big|_{\mathbf{a}}$ is identically 0.
- To classify any of critical points $\mathbf{a}$ into minima, maxima, and saddle points, we looked at the Hessian of f: i.e. $H(f)\big|_{\mathbf{a}}$. Specifically, we said that if the matrix

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix}$$

  was **positive-definite**[1], then $\mathbf{a}$ was a local minimum; similarly, if this matrix is negative-definite, then $\mathbf{a}$ is a local maximum, and if this matrix is nonzero but neither positve nor negative-definite, then $\mathbf{a}$ is a saddle point.

Often, however, we won't just be looking to find the maximum of some function on all of $\mathbb{R}^n$: sometimes, we'll want to maximize a function *given a set of constraints*. For example, we might want to maximize the function $f(x, y, z) = x + y$ subject to the constraint that we're looking at points where $x^2 + y^2 = 1$. How can we do this?

Initially, you might be tempted to just try to use our earlier methods: i.e. look for places where Df is 0, and try to classify these extrema. The problem with this method, when we have a set of constraints, is that it usually **won't** find the maxima or minima on this constraint: because it's only looking for local maxima or minima over all of $\mathbb{R}^n$, it will ignore points that could be maxima or minima on our constrained surface! I.e. for the f, g we mentioned above, we know that $\nabla(f) = (1, 1)$, which is never 0; however, we can easily see by graphing that $f(x, y) = x + y$ should have a maximum value on the set $x^2 + y^2 = 1$, specifically at $x = y = \frac{1}{\sqrt{2}}$.

[1] A $n \times n$ matrix is called **positive-definite** if it has n positive eigenvalues; similarly, it is called **negative-definite** if it has n negative eigenvalues.

So: how can we find these maxima and minima in general? The answer is the method of Lagrange multipliers, which we outline here:

Proposition 1 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function whose extremal values $\{\mathbf{x}\}$ we would like to find, given the constraints $g(\mathbf{x}) = c$, for some constraining function $g(\mathbf{x})$. Then, we have the following result: if $\mathbf{a}$ is an extremal value of f restricted to the set $S = \{\mathbf{x} : \forall i, g(\mathbf{x}) = c\}$, then either one of $\nabla(f)\big|_{\mathbf{a}}, \nabla(g)\big|_{\mathbf{a}}$ are 0, don't exist, or there is some constant λ such that*

$$\nabla(f)\big|_{\mathbf{a}} = \lambda \nabla(g)\big|_{\mathbf{a}}.$$

Why? In this case, it's worth talking a little bit about why this result happens to work, as understanding the proof of the above proposition is remarkably useful for using it. Consider, again, the example we discussed earlier, where we have

$$\begin{aligned} f(x,y) = x + y \qquad & \leftarrow \text{(the function we would like to maximize)}, \\ g(x,y) = x^2 + y^2 = 1 \qquad & \leftarrow \text{(our constraining function)}. \end{aligned}$$

Let

$$S = \{(x,y) : g(x,y) = 1\}.$$

In this notation, we are looking to maximize the function f restricted to the set S, which we denote $f\big|_S$. What do we know about $f\big|_S$? Well: if $\mathbf{a} \in S$ is a maximum, we would expect $\mathbf{a}$ to be a "critical point" of $f\big|_S$. The only issue is that we don't have any way to easily refer to just $f\big|_S$: we can talk about f in general, but if we don't restrict it to S we wouldn't expect $\mathbf{a}$ to still be a maximum.

One way around this is to think about **paths**. Specifically, pick any path γ such that γ's image is constained entirely within S, and $\gamma(0) = \mathbf{a}$. Then, if we look at $f \circ \gamma$, we know that this is a function from $\mathbb{R} \to \mathbb{R}$; therefore, if $f\big|_S$ has a maximum at $\mathbf{a}$, $f \circ \gamma$ *also* must have a maximum, as it's just a path contained entirely in S that goes through this supposed maximum point $\mathbf{a}$.

In other words, we have

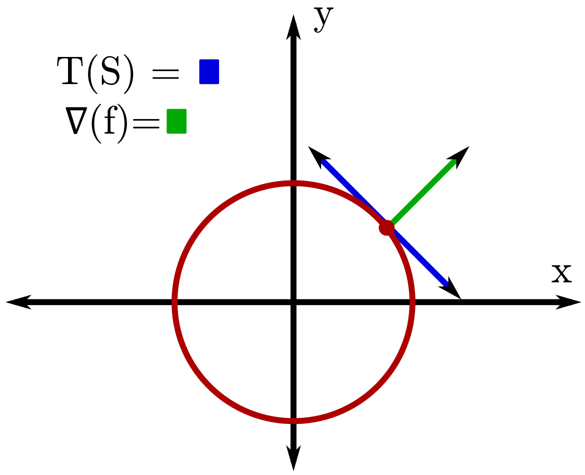
$$\begin{aligned} & \nabla(f \circ \gamma)\big|_{t=0} = 0 \\ \Rightarrow \quad & \nabla(f)\big|_{\mathbf{a}} \cdot \gamma'(0) = 0; \end{aligned}$$

i.e. $\nabla(f)\big|_{\mathbf{a}}$ is orthogonal to $\gamma'(0)$, for any path γ in S, going through 0. But these $\gamma'(0)$'s are just all of the possible tangent vectors to S at $\mathbf{a}$: so we have that $\nabla(f)\big|_{\mathbf{a}}$ is orthogonal to all of these tangent vectors!

Similarly, we know that for any such path γ, we have that $g \circ \gamma$ is constant, because g is constant on S. But this means that (because the derivative of any constant is 0)

$$\begin{aligned} & \nabla(g \circ \gamma)\big|_{t=0} = 0 \\ \Rightarrow \quad & \nabla(g)\big|_{\mathbf{a}} \cdot \gamma'(0) = 0. \end{aligned}$$

In other words, $\nabla(g)\big|_{\mathbf{a}}$ is *also* orthogonal to all of S's tangent vectors at $\mathbf{a}$!



But S is a space formed by placing one constraint on a function of n variables: in other words, it's a $n-1$-dimensional space! Therefore, at the point $\mathbf{a}$, the collection of tangent vectors to S at $\mathbf{a}$ is a $n-1$-dimensional space, contained in $\mathbb{R}^n$. But this means that the space of all vectors orthogonal to this $(n-1)$-dimensional space is a 1-dimensional space! In specific, we've just shown that both $\nabla(f)\big|_{\mathbf{a}}$ and $\nabla(g)\big|_{\mathbf{a}}$ are contained in the same 1-dimensional space: i.e. that one of them is a multiple of the other! In other words, we've shown that because they're both orthogonal to the entire tangent space to S at $\mathbf{a}$, there is some λ such that

$$\nabla(f)\big|_{\mathbf{a}} = \lambda \nabla(g)\big|_{\mathbf{a}}$$

(Or one of them is 0, or undefined.)

In the very specific case we're working with where

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

we have

$$\nabla(f(x,y)) = (1,1), \nabla(g(x,y)) = (2x, 2y)$$

and we're looking for points where either of these gradients are 0, or where there is some λ such that

$$\begin{aligned}
&\nabla(f(x,y)) = (1,1) = \lambda \cdot \nabla(g(x,y)) = (2\lambda x, \lambda y) \\
\Rightarrow &1 = 2\lambda x,\ 1 = 2\lambda y \\
\Rightarrow &\frac{1}{2\lambda} = x,\ \frac{1}{2\lambda} = y \\
\Rightarrow &x = y.
\end{aligned}$$

So: we have either $(0,0)$, as this forces $(2x, 2y) = (0,0)$, or points (x,y) where $x = y$. The first is impossible if we're looking at points where $g(x,y) = x^2 + y^2 = 1$; for the second, we would have $x^2 + x^2 = 1$, i.e .$x = y = \pm\frac{1}{\sqrt{2}}$.

We've therefore discovered the two possible extremal points of f: $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. In other words, we know that these points are the possible local maxima and minima of f. How do we tell whether these points are actually **global** minima and maxima? The answer is in the following brief definitions and theorem:

Definition. A set $S \subset \mathbb{R}^n$ is called **closed** if it contains all of its limit points: i.e. if $\{x_n\}_{n=1}^{\infty} \subset S$,, and $\lim_{n\to\infty} x_n = L$, then $L \in S$.

Definition. A set $S \subset \mathbb{R}^n$ is called **bounded** if there is some M such that $||\mathbf{x}|| < M$, for every $\mathbf{x} \in S$.

Lemma 2 *If $g : \mathbb{R}^n \to \mathbb{R}$ is a continuous function and c is any constant, then the set $S = \{\mathbf{x} : g(\mathbf{x}) = c\}$ is closed.*

Theorem 3 *If f is a continuous function and we restrict f to a closed and bounded set S, then $f\big|_S$ will hit its global minima and maxima on S, and furthermore do this at critical points: i.e. places where $D(f\big|_S)$ is 0.*

Corollary 4 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is some function we want to maximize/minimize, $g : \mathbb{R}^n \to \mathbb{R}$ is some constraint function, S is the constrained set $\{\mathbf{x} : g(\mathbf{x}) = c\}$, and S is a bounded set. Then the absolute maxima and minima of g can all be found via the method of Lagrange multipliers: i.e. the maxima and minima of $f\big|_S$ will come up in the extremal points that the method of Lagrange multipliers finds.*

As a result of this theorem, we know that in our example earlier, one of the two points $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ is the maximum of f, while the other must be its minimum. By plugging in both values to f, we can see that the first is the maximum, and the second is the minimum.

To illustrate the power and versatility of the method of Lagrange multipliers, and to help you get a better feel for how they work in practice, we work two examples using the tools we've just developed:

Example. Consider the **astroid**, a curve in $\mathbb{R}^2$ formed by the equation

$$x^{2/3} + y^{2/3} = 1.$$

What points on this curve are the closest to the origin?

Solution. We want to minimize the distance function

$$f(x, y) = \sqrt{x^2 + y^2}$$

given the constraint

$$g(x, y) = x^{2/3} + y^{2/3} = 1.$$

Using the method of Lagrange multipliers, we know that these minimal points will be those for which either $\nabla(f)$ or $\nabla(g)$ are undefined, or such that there is some λ such that

$$\nabla(f)\big|_{\mathbf{a}} = \lambda\nabla(g)\big|_{\mathbf{a}}.$$

So: calculating, we can see that

$$\nabla(f) = \left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}}\right),$$

which is defined whenever $(x, y) \neq (0, 0)$, and

$$\nabla(g) = \left(\frac{2}{3}\cdot x{-}1/3, \frac{2}{3}\cdot y{-}1/3, \right),$$

which is defined whenever $x \neq 0$ and $y \neq 0$.

When either x or $y = 0$, we know that (in order to satisfy $x^{2/3}+y^{2/3} = 1$) the other value has to be ± 1; so we immediately know that we should look at the four points $(\pm 1, 0), (0, \pm 1)$ when we're looking for our extremal points. Apart from these locations, we know that both of these gradients are well-defined and nonzero; so we're looking for values x, y, λ such that

$$\nabla(f) = \left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}}\right) = \lambda\nabla(g) = \left(\frac{2\lambda}{3}\cdot x{-}1/3, \frac{2\lambda}{3}\cdot y{-}1/3, \right).$$

By equating these two vectors, we're just trying to solve the two equations

$$\begin{aligned}
\frac{x}{\sqrt{x^2+y^2}} &= \frac{2\lambda}{3}\cdot x{-}1/3,\\
\frac{y}{\sqrt{x^2+y^2}} &= \frac{2\lambda}{3}\cdot y{-}1/3\\
\Rightarrow \frac{1}{\sqrt{x^2+y^2}} &= \frac{2\lambda}{3}\cdot x{-}4/3,\\
\frac{1}{\sqrt{x^2+y^2}} &= \frac{2\lambda}{3}\cdot y{-}4/3\\
\Rightarrow x^{-4/3} &= y^{-4/3}\\
\Rightarrow x &= \pm y.
\end{aligned}$$

The only points that satisfy $x = \pm y$ and also $x^{2/3} + y^{2/3}$ are the four points $\left(\pm\frac{1}{2\sqrt{2}}, \pm\frac{1}{2\sqrt{2}}\right)$. Combining these with the $(\pm 1, 0), (0, \pm 1)$ points we discovered earlier, we have eight possible extremal points. Plugging these into $f(x)$ gives us $\frac{1}{2}$ for the points with $x = \pm y$ and 1 for the points with one of $x, y = 0$. Because the maximum and minimum values of f occur on these points, we know that the closest points to the origin are precisely the points $\left(\pm\frac{1}{2\sqrt{2}}, \pm\frac{1}{2\sqrt{2}}\right)$.

Example. Consider the following rough model for the economics of pie-baking:

- Your ingredients for a pie are apples (a), butter (b), flour (f), and sugar (s).
- Suppose that apples cost \$2/unit, butter costs \$3/unit, flour costs \$1/unit, and sugar costs \$1/unit.
- Suppose that if you have a units of apples, b units of butter, f units of flour, and s units of sugar, you can make roughly $\sqrt[4]{abfs}$- many pies. (This is not an entirely implausible guess for a function that tells you how many pies you can make: in particular, you want a function that (1) is 0 whenever you don't have one of your ingredients, which taking the product of all of your ingredients does for you, and (2) grows linearly if you increase the quantity of each of your ingredients linearly. [i.e. if you have k units of each ingredient, this says you can make k pies, which seems accurate.] The formula also allows you to slightly skew the ingredient proportions of your pies: if apples are really expensive, you can have pies that have more dough to apples, whereas if sugar gets really expensive you can just increase the apple ratio.)
- Finally, suppose you start with 100 units of currency, and that you cannot have a negative amount of any of our ingredients (i.e. $a, b, f, s \geq 1$.)

What is the maximum number of pies you can make?

Solution. This is a bit different than our earlier problems. In particular, instead of just optimizing the function

$$F(a, b, f, s) = \sqrt[4]{abfs}$$

on one constraint, we are optimizing it over the **inequalities**

$$2a + 3b + f + s \leq 100, a \geq 0, b \geq 0, f \geq 0, s \geq 0.$$

How can we do this with Lagrange multipliers? Well: to do this, we just need to consider several cases. Specifically, suppose we have some maximum point (a, b, f, s). There are two possibilities:

1. This maximum point occurs on the interior of the set formed by our constraints $2a + 3b + f + s \leq 100, a \geq 0, b \geq 0, f \geq 0, s \geq 0$. Therefore, this point can be found by looking at $D(f)$, as it's a local maximum of f without any constraints!
2. Otherwise, this maximum point occurs on the boundary of the set formed by the constraints $2a + 3b + f + s \leq 100, a \geq 0, b \geq 0, f \geq 0, s \geq 0$. In other words, this maximum point occurs when we have either

   $$g(a, b, f, s) = 2a + 3b + f + s = 100,$$

   or when one of the four quantities a, b, f, s are 0. We can eliminate the cases where we have 0 of any of our quantity by just noticing that this trivially restricts us to making 0 pies, which is clearly not a maximum; this leaves us with just the above single constraint. But this is exactly the situation that Lagrange multipliers are set up to deal with! In particular, we can use Lagrange multipliers to maximize $\sqrt[4]{abfs}$ with respect to the constraint $g(a, b, f, s) = 2a + 3b + f + s = 100$.

Comparing all of the critical points we find in these ways will yield the overall maximum of $\sqrt[4]{abfs}$ on our entire set.

We perform these calculations here. First, because

$$\nabla(F) = \left(\frac{bfs}{4(abfs)^{3/4}}, \frac{afs}{4(abfs)^{3/4}}, \frac{abs}{4(abfs)^{3/4}}, \frac{abf}{4(abfs)^{3/4}} \right),$$

we can see that the gradient of F is only undefined or zero at places where some of its coördinates are zero or negative. Because our conditions require that $a, b, f, s \geq 0$, and in the case that any quantity is 0 we know that no pies are made, we know that this is impossible: therefore, we don't have to worry about f having any maxima on the interior of our set of constraints.

Now, we turn to the constraint $g(a, b, f, s) = 2a + 3b + f + s = 100$. Using Lagrange multipliers, we know that critical points will occur where

$$\nabla(F) = \left(\frac{bfs}{4(abfs)^{3/4}}, \frac{afs}{4(abfs)^{3/4}}, \frac{abs}{4(abfs)^{3/4}}, \frac{abf}{4(abfs)^{3/4}} \right) = \lambda \nabla(g) = (2\lambda, 3\lambda, \lambda, \lambda).$$

This occurs at points that satisfy the four equations

$$\frac{bfs}{4(abfs)^{3/4}} = 2\lambda, \frac{afs}{4(abfs)^{3/4}} = 3\lambda, \frac{abs}{4(abfs)^{3/4}} = \lambda, \frac{abf}{4(abfs)^{3/4}} = \lambda.$$

By combining the first two equations, we can see that

$$\begin{aligned} &\frac{bfs}{4(abfs)^{3/4}} = 2\lambda, \frac{afs}{4(abfs)^{3/4}} = 3\lambda \\ \Rightarrow &b = 2\left(4\lambda \frac{1}{fs}(abfs)^{3/4}\right), a = 3\left(4\lambda \frac{1}{fs}(abfs)^{3/4}\right) \\ \Rightarrow &\frac{b}{2} = \frac{a}{3}. \end{aligned}$$

Similarly, we can combine the middle two equations to get

$$\begin{aligned} &\frac{afs}{4(abfs)^{3/4}} = 3\lambda, \frac{abs}{4(abfs)^{3/4}} = \lambda \\ \Rightarrow &f = 3\left(4\lambda \frac{1}{as}(abfs)^{3/4}\right), b = \left(4\lambda \frac{1}{as}(abfs)^{3/4}\right) \\ \Rightarrow &\frac{f}{3} = b, \end{aligned}$$

and(similarly) the last two equations to get $f = s$. Combining these results, we can write our point as $(a, \frac{2a}{3}, 2a, 2a)$, and get that

$$\begin{aligned} &2a + 3b + f + s = 2a + 2a + 2a + 2a = 100 \\ \Rightarrow &a = 12.5, (a, \frac{2a}{3}, 2a, 2a) = (12.5, 8.\overline{3}, 25, 25). \end{aligned}$$

With these ingredient ratios, we can make

$$\sqrt[4]{12.5 \cdot 8.\overline{3} \cdot 25 \cdot 25)} \cong 16$$

pies; as this is the only critical point that gives a nonzero value, we know that it must be our maximum.