Math 1540
Spring 2011
Notes #7
More from chapter 7

1 An example of the implicit function theorem

First I will discuss exercise 4 on page 439. The problem is to say what you can about solving the equations

$$x^2 - y^2 - u^3 + v^2 + 4 = 0 \tag{1}$$
$$2xy + y^2 - 2u^2 + 3v^4 + 8 = 0 \tag{2}$$

for u and v in terms of x and y in a neighborhood of the solution $(x, y, u, v) = (2, -1, 2, 1)$.

Let

$$F(x, y, u, v) = \left(x^2 - y^2 - u^3 + v^2 + 4, 2xy + y^2 - 2u^2 + 3v^4 + 8\right),$$

so that $F(2, -1, 2, 1) = (0, 0)$. The implicit function theorem says to consider the Jacobian matrix with respect to u and v. (You always consider the matrix with respect to the variables you want to solve for. This is obvious in the one-dimensional case: if you have $f(x, y) = 0$ and you want y to be a function of x, then you differentiate $f(x, y) = 0$ with respect to x, getting

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0,$$

or

$$\frac{dy}{dx} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}.$$

It is $\frac{\partial f}{\partial y}$ which must not be zero.)

In our two dimensional case, therefore, we consider

$$A = \begin{pmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{pmatrix} |_{(2,-1,2,1)} = \begin{pmatrix} -12 & 2 \\ -8 & 12 \end{pmatrix},$$

so

$$\Delta = -128 \neq 0.$$

Hence we can solve for (u,v) in a neighborhood of $(x_0,y_0)=(2,-1)$. More precisely, there are neighborhoods U of $(2,-1)$ and W of $(2,1)$ such that for each $(x,y)\in U$ there is a unique (u,v) in W such that $F(x,y,u,v)=0$. Hence, there is a unique function $f:U\to W$ such that $F(x,y,f_1(x,y),f_2(x,y))=0$ for every $(x,y)\in U$.

Further, the problem asks us to compute $\frac{\partial u}{\partial x}$ in this neighborhood. To do this we differentiate the equations (1)-(2) implicitly with respect to x. We get

$$\begin{aligned}2x-3u^3\frac{\partial u}{\partial x}+2v\frac{\partial v}{\partial x}&=0\\ 2y-4u\frac{\partial u}{\partial x}+12v^3\frac{\partial v}{\partial x}&=0.\end{aligned}$$

This is a set of two equations in the two unknowns $\frac{\partial u}{\partial x}$ and $\frac{\partial v}{\partial x}$. Putting the terms with the unknowns on the other side, and solving for this unknowns, we get

$$\begin{aligned}\begin{pmatrix}\frac{\partial u}{\partial x}\\ \frac{\partial v}{\partial x}\end{pmatrix}&=\begin{pmatrix}3u^2 & -2v\\ 4u & -12v^3\end{pmatrix}^{-1}\begin{pmatrix}2x\\ 2y\end{pmatrix}\\ &=\frac{1}{8uv-36u^2v^2}\begin{pmatrix}-12v^3 & 2v\\ -4u & 3u^2\end{pmatrix}\begin{pmatrix}2x\\ 2y\end{pmatrix}.\end{aligned}$$

Hence,

$$\frac{\partial u}{\partial x}=\frac{1}{8uv-36u^2v^2}\left(-24xv^3+4vy\right).$$

Note that this is a function of both (x,y) and (u,v). To get it entirely in terms of (x,y) would require actually solving for (u,v) in terms of (x,y), and, as in this case, this is often very messy or impossible. But often it is good enough to know get the derivatives only at (x_0,y_0,u_0,v_0). For example, this will give a first order approximation to (u,v) in terms of (x,y) in a neighborhood of (x_0,y_0).

2 The “domain straightening theorem".

This appear to be a rather strange theorem, and indeed, it is mainly useful theoretically. For example, in this chapter it is used in the proof of an important result, the theorem about Lagrange multipliers. It is not usually used directly in an application.

As an example of this theorem in two dimensions, first consider the linear function

$$f(x,y)=2x+y.$$

the goal is to change coordinates in the (x,y) plane to make the function even simpler. To do this, we let

$$(u,v)=G(x,y)=(x,2x+y).$$

Then,

$$(x, y) = G^{-1}(u, v) = (u, v - 2u).$$

If $h = G^{-1}$, then

$$f(h(u, v)) = f(u, v - 2u) = 2u + (v - 2u) = v.$$

The function $f(h(u, v))$ is a very simple function, that is, just the "projection" onto the v axis.

Think of this in terms of the graph of $z = 2x + y$. Introduce new coordinates u and v, with $u = x, v = 2x + y$. Then the surface has the simple equation $z = v$.

We wish to do this for a nonlinear function. Find a nonlinear change of coordinates which simplifies the function to a projection. In the new coordinates, the graph of z will be a plane.

Consider the example

$$f(x, y) = x^2 + y.$$

Let $G(x, y) = (x, x^2 + y)$. Then $G^{-1}(u, v) = (u, v - u^2)$. (Think about it!). Then

$$f\left(G^{-1}(u, v)\right) = f\left(u, v - u^2\right) = u^2 + v - u^2 = v.$$

In (u, v) coordinates, the surface is just the slanted plane $z = v$.

An additional complication is seen by considering $f(x, y) = x + y^2$. If we let $G(x, y) = (x, x + y^2)$, then G^{-1} does not exist, because we can't solve $x = u$, $x + y^2 = v$ for y uniquely in terms of (u, v). This is because $\frac{\partial G}{\partial y}(0, 0) = 0$. The simplest thing is to work with u instead of v. Let $G(x, y) = (x + y^2, y)$. Then the inverse is

$$G^{-1}(u, v) = \left(u - v^2, v\right)$$

The surface $z = x + y^2$ becomes the plane $z = u$. That is, $f(G^{-1}(u, v)) = f(u - v^2, v) = u - v^2 + v^2 = u$.

The book goes through some extra manipulations to find a function $(u, v) = h(x, y)$ such that the surface is still projection onto the second coordinate, giving $f(x, y) = v$. This is not hard; just switch the coordinates, find the function $h = G^{-1}$, and switch them back again, but it becomes more confusing to read. I will not cover this here.

3 Lagrange Multipliers

This is an important maximization (or minimization) technique.

Theorem 1 *Suppose that* $f : R^n \to R$ *and* $g : R^n \to R$ *are* C^1 *functions. Suppose that for some* $c_0 \in R$ *there is an* $x_0 \in R^n$ *such that* $g(x_0) = c_0$. *Let* $S = g^{-1}(c_0)$. *(In any interesting case,* g *is not* $1:1$, *and* $g^{-1}(c_0)$ *contains at least a curve in* R^n *through* x_0.*) Suppose that* $f|_S$ *has a local maximum or local minimum at some point* x_1, *and that* $Dg(x_1) \neq 0$. *(*$Dg(x_1) = \nabla g(x_1)$.*) then there is a scalar* λ *such that*

$$\nabla f(x_1) = \lambda \nabla g(x_1). \tag{3}$$

Remark 2 *In the text, the statement assumes that the original point* x_0 *is the maximum;* $x_1 = x_0$. *In practice you are not given* x_0. *(Otherwise you would be done!) You are given a constraint* $g = c_0$, *and so in theory you know the set* $S = g^{-1}(c_0)$. *You can usually see immediately that* S *is not empty, meaning that there is an* x_0 *with* $g(x_0) = c_0$. *You look for a point on* S *where* f *is a maximum. (If* S *is not compact, then you have to be concerned about whether there is any maximum.)*

Remark 3 *Notice that equation (3) is a* ***necessary*** *condition for a max or min, not a sufficient condition. It is a type of first derivative test, and first derivative tests (*$f' = 0$*) are never sufficient, just necessary. If* S *is compact, then you know that there must be a global max and a global min of* $f|S$.[1] *To find it you can examine all the points where (3) is satisfied, and see at which* f *is the largest. In many cases, you will find only two points. Then one must be the maximum and the other the minimum.*

Proof. We start with a definition.

Definition 4 *The "tangent space" to* S *at a point* $x_1 \in S$ *is the subspace of* R^n *which is orthogonal to* $\nabla g(x_1)$.

(This is often denoted by $[\text{span}\{\nabla g(x_1)\}]^\perp$. You need to remember some linear algebra. We know that $\nabla g(x_1)$ is a vector in R^n, and we have assumed that it is not the zero vector. The set

$$V = \{v \in R^n \mid v \cdot \nabla g(x_1) = 0\}$$

is easily shown to be an $n-1$ dimensional subspace of R^n. If $W = \text{span}\{\nabla g(x_1)\}$, then $W^\perp = V$.)

[1] The text does not appear to use the terms global max and global min. I just mean the maximum and the minimum; the word global is inserted simply to emphasize that these are more than just local.

Suppose that $c : R \to R^n$ is a smooth (continuously differentiable) curve lying in S, with $c(0) = x_1$. Then $f \circ c : R \to R$, and $f \circ c$ has a local maximum at 0.Hence, $Df(x_1)\, c'(0) = 0$.[2] Hence, $\nabla f(x_1)$ is perpendicular to $c'(0)$. Also, $c'(0)$ is tangent to S, and so lies in the tangent space V. Thus, $\nabla g(x_1)$ is also perpendicular to $c'(0)$. But this is not enough to show that $\nabla g(x_1)$ is in the same direction as $\nabla f(x_1)$, which is what (3) implies.

However, $\nabla f(x_1)$ is perpendicular to $c'(0)$ for **every** smooth curve passing through c_0. So it seems very likely that $\nabla f(x_1)$ is a scalar multiple of $\nabla g(x_1)$, which is what (3) implies. To prove this we need to know that $\nabla f(x_1)$ is perpendicular to every vector in V. Then $\nabla f(x_1) \in V^{\perp}$, and so $\nabla f(x_1) \in [\text{span}\{\nabla g(x_1)\}]$, meaning that $\nabla f(x_1)$ is a scalar multiple of $\nabla g(x_1)$.

We have not shown that for each vector $v \in V$ there is a smooth curve c for which $c'(0)$ is parallel to v. The proof of this (pg. 434) uses the domain straightening theorem, but I will leave this for you to read. ■

One reason I omit this last detail is that in the applications we will discuss, S is either a curve in R^2, in which case there is only one tangent line, or it is a smooth surface in R^3, when the result is also pretty "obvious". Example 7.7.5 is of the first type, and examples 7.7.6 and 7.7.7 are of the second type. I will go over some of these in class.

4 Homework, due Feb. 23.

1. pg. 443, # 30
2. pg. 439, #5 a,c.
3. pg,. 443, # 36.

There will also be one or two exercises on material in the next set of notes, which are not taken from the text. They are laying the groundwork for Stokes' theorem, a topic not in the text.

[2] Recall from section 6.3, where smooth curves were considered, that $Dc(t) = c'(t) = (c_1'(t), ..., c_n'(t))$. This is the tangent vector to the curve (actually, the image of the curve, since the curve is the function c,)