INVERSE/IMPLICIT FUNCTION THEOREM AND LAGRANGE MULTIPLIERS

Theorem 1. *[Inverse Function Theorem] Suppose* $F:\mathbb{R}^n \to \mathbb{R}^n, F = (F_1(x_1,x_2,\dots x_n), F_2(x_1,x_2,\dots x_n),\dots F_n(x_1,x_2,\dots x_n))$ *is continuously differentiable and has* $\det \frac{\partial(F_1,F_2,\dots,F_n)}{\partial(x_1,x_2,\dots,x_n)} \neq 0$ *at some point* a. *Then there is an open n'h'd* U *of* a *so that* $F:U\to F(U)$ *is invertible. Moreover, the inverse function* $F^{-1}:F(U)\to U$ *is also continuously differentiable and satisfies:*

$$\frac{\partial(F_1,F_2,\dots,F_n)}{\partial(x_1,x_2,\dots,x_n)}(a) = \left[\frac{\partial(F_1,F_2,\dots,F_n)}{\partial(x_1,x_2,\dots,x_n)}\right]^{-1}(F(a))$$

Proof. [sketch] Can assume WOLOG (by applying a linear transformation to F) that $\frac{\partial(F_1,F_2,\dots,F_n)}{\partial(x_1,x_2,\dots,x_n)}(a) = Id$. Now, by the continuity of derivatives, find a n'h'd U of a so that $\left|\frac{\partial F_j}{\partial x_k}(x) - \frac{\partial F_j}{\partial x_k}(a)\right|$ is small in the entire n'h'd. Use this, and integration inequalitys, to argue that F must be 1:1 in U. Some more work gives that the inverse has to be continuous and differentiable. Once that is established, use chain rule to get the rules for the derivative of the inverse. □

Theorem 2. *[Implicit Function Theorem] Let* $f:\mathbb{R}^{n+k}\to\mathbb{R}^n$. *We think of* $\mathbb{R}^{n+k}$ *as* $\mathbb{R}^n\times\mathbb{R}^k$ *here, and we write*$f(\vec{x},\vec{y}) = (f_1(x_1\dots x_n,y_1,\dots y_k),\dots f_n(x_1\dots x_n,y_1,\dots y_k))$. *If* f *is continuously differentiable, and there is some point* (a,b) *with* $f(a,b)=\vec{0}$ *and if the following Jacobian matrix is non-zero:*

$$\det\left[\frac{\partial(f_1,\dots f_n)}{\partial\,(x_1,\dots x_n)}(a,b)\right]\neq 0$$

Then there is a continuously differentiable function $h:\mathbb{R}^k\to\mathbb{R}^n$ *defined in a n'h'd of* a *so that the x-coordinates can be written as an implicit function of the y-coordinates:*

$$\left\{(x,y): f(x,y)=\vec{0}\right\} = \{(h(y),y)\}$$

Proof. (Proof assuming the Inverse Function Theorem) Define $F:\mathbb{R}^{n+k}\to\mathbb{R}^{n+k}$ by padding f with the identity: $F(\vec{x},\vec{y}) = (f(\vec{x},\vec{y}),\vec{y})$ Notice that the Jacobian matrix of F is block diagonal:

$$\frac{\partial(F_1,\dots F_n,F_{n+1},\dots F_{n+k})}{\partial\,(x_1,\dots x_n,y_1\dots y_k)} = \begin{bmatrix} \frac{\partial(f_1,\dots f_n)}{\partial(x_1,\dots x_n)} & \frac{\partial(f_1,\dots f_n)}{\partial(y_1,\dots,y_k)} \\ 0_{k\times n} & Id_{k\times k}\end{bmatrix}$$

So we have that $\det\left(\frac{\partial F}{\partial(\vec{x},\vec{y})}\right)(a,b) = \det\left(\frac{\partial f}{\partial \vec{x}}\right)(a,b)$. When this is non-zero, we know by application of the inverse function theorem that we can find a smooth inverse function in a n'h'd of (a,b). In particular, if we look in the n'h'd of the point $\left(\vec{0},\vec{b}\right)$ we find some inverse function F^{-1}. Denote the first k components of F^{-1} by the function $h:\mathbb{R}^k\to\mathbb{R}^n$ defined by:

$$F^{-1}\left(\vec{0},\vec{y}\right) = (h(\vec{y}),\vec{y})$$

[Note that we know the last k components of F^{-1} here are the identity by inspecting of the definition of F] Notice that because F^{-1} is an inverse function to F in a n'h'd of (a,b) we have that:

$$\begin{aligned}\left(\vec{0},\vec{y}\right) &= F\left(h(\vec{y}),\vec{y}\right)\\ &= \left(f(h(\vec{y}),\vec{y}),\vec{y}\right)\end{aligned}$$

In other words:

$$f(h(\vec{y}),\vec{y}) = \vec{0}$$

The equality of these sets in the statement of the theorem follows because F is locally a bijection. We can also get information about derivatives from the statement about the derivatives from the inverse function theorem. □

Example 3. (Sept 03 #5) Suppose that $F=F(u,v)$ is a smooth function from $\mathbb{R}^2\to\mathbb{R}$ with $F(1,2)=0$ and $F_v(1,2)\neq 0$. Show that:

$$F(xy,\sqrt{x^2+z^2}) = 0$$

defines a smooth surface $(x,y,z(x,y))$ in a n'h'd of $(1,1,\sqrt{3})$.Find a normal vector $n\neq 0$ to this.

Theorem 4. *[Lagrange Multipliers] Suppose we are interested in optimizing (finding the maximum or minimum) of $f(x,y)$ which is continuously differentiable subject to the constraint that $g(x,y)=c$ which is continuously differentiable. IF (x_0,y_0) is a local maximum for this problem, THEN there exists $\lambda_0$so that:*

$$\nabla f(x_0,y_0)=\lambda_0\nabla g(x_0,y_0)$$

Proof. [Proof using ∇g is perpendicular to $g(x,y)=c$] Let $r(t)=(x(t),y(t))$ be any curve in the surface $g(x,y)=c$ with $r(0)=(x_0,y_0)$. Since x_0,y_0 is a local maximum, we know that $\frac{d}{dt}f(r(t))|_{t=0}=0$. By the chain rule, we know then that $\nabla f(x_0,y_0)\cdot r'(0)=0$. This works for *any curve* $r(t)$ though, so it must be that $\nabla f(x_0,y_0)$ is perpendicular to the surface $g(x,y)=c$ at the point (x_0,y_0). But we know that $\nabla g(x_0,y_0)$ is too, so they must be parallel. □

Remark 5. Be careful of the statement of the theorem. It says IF its a local extremum, then $\nabla f=\lambda\nabla g$. In practice, you find a set of candidate points where $\nabla f=\lambda\nabla g$ and then investigate them to see which are the true extrema.

Remark 6. You can make a more careful proof using the implicit function theorem. There is also a more "picture" proof that is described in September 2006 #4 using the function $\phi(x,y,\lambda)=f(x,y)-\lambda(g(x,y)-c)$

Remark 7. In general, for more than one constraint, we have that the set of vectors:

$$\{\nabla f,\nabla g_1,\nabla g_2,\ldots,\nabla g_k\}$$

Must be dependent.

Example 8. (Jan 09 #3) Let $\Sigma=\{x\in\mathbb{R}^3: x_1x_2+x_1x_3+x_2x_3=1\}$ and $f(x)=x_1^2+x_2^2+\frac{9}{2}x_3^2$. Show that Σ is a smooth surface in $\mathbb{R}^3$. Show that $\inf_{x\in\Sigma}f(x)$ is achieved. Find $\inf_{x\in\Sigma}f(x)$.

Example 9. (Jan 10 #3) Let x,y,z be positive real numbers. Show that the following inequality holds (Hint: Use homogeneity):

$$\frac{1}{x^3}+\frac{16}{y^3}+\frac{1}{z^3}\geq\frac{256}{(x+y+z)^3}$$

Example 10. (Sept 06 #4) Part 1: Explain the idea behind the method of Lagrange multipliers for finding an extremum of$f(x,y)$ subject to the constraint $g(x,y)=a$ with $a\in\mathbb{R}$ and smooth f and g. (Please use the function $\phi(x,y,\lambda)=f-\lambda(g-a)$).

Part 2: Show that $df*/da=\lambda*$ where $f*(a)$ is the value of f at the conditional extremum and $\lambda*$ is the corresponding value of the Lagrange multiplier.

Part 3: Find the minimum of x^2+y^2 subject to $x-y=1$. Without recomputing, what is your best estimate for the minimum if the constraint is changed to $x-y=2$ or $x-y=0$?

1. Some Topics I didn't cover

- Intermediate/Mean Value Theorems (I think I used these a few times but never specifically mentioned them. They are some of my favourite theorems, you probably know them already! They might be used as an ingredient in other problems)
- Fourier Series (There is a whole theory of Fourier series that could fill an entire class. You should decide if you want to learn some for the writtens. Usually on the writtens, you don't have to be very careful with Fourier series, so the real problems just become manipulating integrals/sums.)
- Fubini/Tonelli (Just says that for positive $g\geq 0$ that $\int\int g(x,y)dxdy=\int\int g(x,y)dydx$ and for arbitrary f $\int\int|f(x,y)|dxdy<\infty$ then $\int\int f(x,y)dxdy=\int\int f(x,y)dydx$....similar to absolutely summable.)
- Functional Relations (These have a very "one-of" flavor...every problem is different. The only trick that is sometimes useful is to show something for rationals, and then say what it must be everywhere since rationals are dense in $\mathbb{R}$)