APPLICATIONS OF THE INVERSE AND IMPLICIT FUNCTION THEOREM

DR. SARAH RAYNOR

In these notes we will discuss examples of the Inverse and Implicit Function Theorems, which were partially covered in our abbreviated class on Wednesday, April 25. These notes cover some of the main uses of these theorems in a standard course on multivariable calculus.

1. Level Sets of Multivariate Functions

Suppose that $f : \mathbb{R}^{n+1} \to \mathbb{R}$. We will express elements of $\mathbb{R}^{n+1}$ using the notation $(z, \vec{x})$, where $\vec{x} \in \mathbb{R}^n$. Suppose that $(z_0, \vec{x}_0) \in \mathbb{R}^{n+1}$ such that $F(z_0, \vec{x}_0) = 0$. Then the Implicit Function Theorem states that if $\frac{\partial f}{\partial z}(z_0, \vec{x}_0) \neq 0$, then there exist neighborhoods $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}$ so that $\vec{x}_0 \in U$ and $z_0 \in V$ and a function $g : U \to V$ so that, for any $\vec{x} \in U$, $f(g(\vec{x}), \vec{x}) = 0$. That is, the level set of the function f, which is defined to be $f^{-1}(0) = \{(z, \vec{x}) \in \mathbb{R}^{n+1} : f(z, \vec{x}) = 0\}$, can be expressed as the graph of the function $z = g(\vec{x})$, at least locally near the point $(z_0, \vec{x}_0)$. Another way to say this is that the surface defined by $F(z, \vec{x}) = 0$ may not be the graph of a function globally in all of $\mathbb{R}^{n+1}$, and it may not be possible at all to solve explicitly for z as a function of $\vec{x}$, but it is *usually* possible to implicitly express z as a function of $\vec{x}$ is small neighborhoods.

Moreover, the Implicit Function Theorem also states that at any such point where $\frac{\partial f}{\partial z} \neq 0$, then $\nabla_{\vec{x}} g(\vec{x}) = \frac{-1}{\frac{\partial f}{\partial z}} \nabla_{\vec{x}} f(z, \vec{x})$.

Let's look at how this works in a simple example:

Example 1.1. *Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(z, x) = z^2 + x^2 - 1$. Then the set where $f(z, x) = 0$ is the unit circle $z^2 + x^2 = 1$. Clearly, this set is not, in its entirety, the graph of a function, because it does not pass the vertical line test. However, either the upper or lower semicircle is a function. Therefore, if we start at a point (z_0, x_0), then the level curve is a function in a neighborhood of (z_0, x_0) unless $(z_0, x_0) = (0, 1)$ or $(0, -1)$–the two points where the graph has a vertical tangent line and then turns back on itself. In this case, the equation is sufficiently simple that we can actually solve for z in terms of x: in the upper semicircle we have $z = \sqrt{1 - x^2}$, and in the lower semicircle we have $z = -\sqrt{1 - x^2}$. In both cases it is possible to check (as discussed in class), that $\frac{dz}{dx} = -\frac{x}{z}$.*

Now let's look at this from the perspective of the Implicit Function Theorem. We expect to be able to solve for z in terms of x as long as $\frac{\partial f}{\partial z} \neq 0$. We have that $\frac{\partial f}{\partial z} = 2z \neq 0$ so long as $z \neq 0$. Hence this is possible except at the points $(0, 1)$ and $(0, -1)$, exactly as found above. Moreover, in this case we expect that $\frac{dg}{dx} = \frac{-1}{\frac{\partial f}{\partial z}} \frac{\partial f}{\partial x} = -\frac{x}{z}$, again as computed above. Therefore, our direct calculations agree with the conclusions of the implicit function theorem.

As the dimension increases, we see a little more complication, but the ideas are fundamentally the same.

Date: Wednesday, April 25, 2012.

Example 1.2. *Suppose that $f : \mathbb{R}^3 \to \mathbb{R}$ is given by $f(z, x) = z^2 + x^2 + y^2 - 9$. Then the set where $f(z, x, y) = 0$ is the sphere $z^2 + x^2 + y^2 = 9$. This again does not pass the vertical line test. However, either the upper or lower hemisphere is a function. Therefore, if we start at a point (z_0, x_0, y_0), then the level surface is a function in a neighborhood of (z_0, x_0, y_0) unless $z_0 = 0$, i.e.$x_0^2 + y_0^2 = 9$–the equator, where the graph has a vertical tangent plane and then turns back on itself. In this case, the equation is sufficiently simple that we can actually solve for z in terms of x: in the upper hemisphere we have $z = \sqrt{9 - x^2 - y^2}$, and in the lower hemisphere we have $z = -\sqrt{9 - x^2 - y^2}$. In both cases it is possible to check (as discussed in class), that $\frac{\partial z}{\partial x} = -\frac{x}{z}$ and $\frac{\partial z}{\partial y} = -\frac{y}{z}$.*

Now let's look at this from the perspective of the Implicit Function Theorem. We expect to be able to solve for z in terms of x and y as long as $\frac{\partial f}{\partial z} \neq 0$. We have that $\frac{\partial f}{\partial z} = 2z \neq 0$ so long as $z \neq 0$. Hence this is possible except at the equator, exactly as discussed above. Moreover, in this case we expect that $\frac{\partial g}{\partial x} = \frac{-1}{\frac{\partial f}{\partial z}} \frac{\partial f}{\partial x} = -\frac{x}{z}$ and $\frac{\partial g}{\partial y} = \frac{-1}{\frac{\partial f}{\partial z}} \frac{\partial f}{\partial y} = -\frac{y}{z}$, again as computed above. Therefore, our direct calculations agree with the conclusions of the implicit function theorem.

Exercise 1.1. *Consider the surface $f(z, x, y) = x^3 + 3y^2 + 8xz^3 - 3z^3 y = 1$. Determine near which points (x_0, y_0, z_0) in $\mathbb{R}^3$ it is possible to express z as function of x and y locally, and justify your work via the Implicit Function Theorem. At such points, find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.*

2. Change of Coordinates

One of the most important uses of the Inverse Function theorem is changes of coordinates. Consider a function $T : \mathbb{R}^n \to \mathbb{R}^n$. Suppose that $(x_1, \ldots, x_n)$ is an element of the domain $\mathbb{R}^n$, and express $T(\vec{x}) = (u_1(\vec{x}), \ldots, u_n(\vec{x}))$. Then we can think of $\vec{x}$ as being the original coordinates in $\mathbb{R}^n$ and $\vec{u}$ as being the new coordinates. T is the function which expresses how one can convert a point from $\vec{x}$ terms into $\vec{u}$ terms. Then the inverse function theorem says that T is invertible, i.e. $\vec{x}$ and $\vec{u}$ are equivalent coordinates for $\mathbb{R}^n$ and we can solve for $\vec{x}$ in terms of $\vec{u}$, if DT is invertible. Here, DT is the $n \times n$ matrix of first derivatives of T, usually called the *Jacobian matrix* of T. Then since DT is a square matrix, by standard linear algebraic results, T is invertible if $det(DT) \neq 0$. The function $det(DT)$, or sometimes $|det(DT)|$ is usually called the *Jacobian* of T. It is the same quantity which appears in the change of variables formula for multiple integrals (that is, u-substitution).

Example 2.1. *Consider polar coordinates on $\mathbb{R}^2$. We can express x and y in terms of r and θ as follows: $x = r\cos(\theta)$, $y = r\sin(\theta)$. Then our transformation T is given by $T(r, \theta) = (r\cos(\theta), r\sin(\theta))$, so*

$$DT = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix},$$

and $det(DT) = r$. It therefore follows that the transformation from polar to Cartesian coordinates is invertible unless $r = 0$, i.e. except at the origin. If $r > 0$, then we obtain the following formula for DT^{-1}:

$$DT^{-1} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\frac{\sin(\theta)}{r} & \frac{\cos(\theta)}{r} \end{pmatrix}.$$

Example 2.2. *We may similarly consider spherical coordinates on $\mathbb{R}^3$. We can express the transformation as follows:*

$$T(\rho, \theta, \phi) = (\rho \cos(\theta) \sin(\phi), \rho \sin(\theta) \sin(\phi), \rho \cos(\phi)),$$

where θ is the horizontal angle from the positive x-axis (longitude), which varies from 0 to 2ϕ, and ϕ is the angle from the z-axis (azimuthal angle; latitude), which varies from 0 to π; ρ is the distance from the origin, which can be any nonnegative real number. So

$$DT = \begin{pmatrix} \cos(\theta)\sin(\phi) & -\rho\cos(\theta)\sin(\phi) & \rho\cos(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) & \rho\cos(\theta)\sin(\phi) & \rho\sin(\theta)\cos(\phi) \\ \cos(\phi) & 0 & -\rho\sin(\phi) \end{pmatrix}.$$

It therefore follows that det$(DT) = -\rho^2 \sin(\phi)$, so T is invertible unless $\rho = 0$ or $\phi = 0$ or 2π. This makes sense–spherical coordinates, just like polar coordinates, are singular at the origin where the angles become undefined. They are also singular at the north and south pole where ϕ makes sense, but θ does not.

Exercise 2.1. *Consider the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(u, v) = (u, uv)$. Draw a picture of the image of the square $[0, 1] \times [0, 1]$ under this transformation. Determine when T is invertible. Find the area element for the transformation T.*

Exercise 2.2. *Consider the transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ which converts cylindrical coordinates to cartesian coordinates. Following the examples above, express T in terms of coordinate functions, find DT and det(DT) and use this information and the inverse function theorem to determine on which parts of $\mathbb{R}^3$ the conversion is invertible.*

3. Lagrange Multipliers

Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ are smooth functions. Suppose that we want to maximize the value of the function f subject to the constraint that g equals a given constant c. That is, we want to find the maximum value of $f(x, y)$ on the level curve given by $g(x, y) = c$. This is a very common problem in applications of calculus. For example, we may want to maximize the volume of a cylindrical can subject to a fixed surface area. In that case, the function f is the volume of the can and the function g is the surface area.

This type of problem is usually solved via a technique known as Lagrange multipliers. The main theorem is as follows:

Theorem 3.1. *Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ are smooth functions. Let $c \in \mathbb{R}$. Suppose that $g(x_0, y_0) = c$ and $f(x_0, y_0) \geq f(x, y)$ for every $(x, y) \in \mathbb{R}^2$ such that $g(x, y) = c$. Then if $\nabla g(x_0, y_0) \neq \vec{0}$, it follows that $\nabla f(x_0, y_0)$ is a scalar multiple of $\nabla g(x_0, y_0)$.*

Remark 3.1. *Notice that this is not an existence theorem–it is certainly possible for no maximum to exist, depending on the shapes of f and g.*

Remark 3.2. *Notice also that this is not a uniqueness theorem–it is possible for this technique to produce critical points which are not maxima or even extrema. Moreover, it is possible for there to be critical points at which $\nabla g = \vec{0}$, which this method will not find.*

Remark 3.3. *The conclusion of this theorem is identical in higher dimensions, but for the purposes of space we will prove it only in two dimensions.*

Proof. Suppose that f, g, c, x_0, and y_0 are as in the hypotheses of the theorem. Consider the set $\{(x, y) \in \mathbb{R}^2 : g(x, y) = c$. We have assumed that $\nabla g \neq \vec{0}$, so without loss of generality we may assume that $\frac{\partial g}{\partial y} \neq 0$. (Otherwise we can use the same argument as below but interchange the roles of x and y.) Then by the Implicit Function Theorem there exists a function $h : \mathbb{R} \to \mathbb{R}$ so that, for all x in some neighborhood U of x_0, $g(x, h(x)) = c$, and moreover $\frac{dh}{dx} = -\frac{\frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}}$ Let $\vec{v}$ be any vector in $\mathbb{R}^2$ for which $\vec{v} \cdot \nabla g(x_0, y_0) = 0$. Now, define the curve $\vec{\gamma}(t)$ by $\vec{\gamma}(t) = (x_0 + tv_1, h(x_0 + tv_1))$. Then, by the chain rule,

$$\frac{dg(\gamma(t))}{dt} = \nabla g \cdot (v_1, \frac{dh}{dx} v_1) = \nabla g \cdot (v_1, v_1 - \frac{\frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}}) = \frac{v_1}{\frac{\partial g}{\partial y}} (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}) \cdot (-\frac{\partial g}{\partial y}, \frac{\partial g}{\partial x} = 0.$$

Hence $g(\gamma(t)) = c$ for all times t. Therefore, if $f(x_0, y_0)$ is the maximum value of f on the constraint set $g = c$, it follows that $f(x_0, y_0)$ is also the maximum value of f along the graph of γ. Therefore, it must be the case that $f(\gamma(t))$ has a critical point at $t = 0$. But $\frac{df(\gamma(t))}{dt} = \nabla f \cdot (v_1, v_2)$ at $t = 0$. So, ∇f is orthogonal to any vector to which ∇g is orthogonal. Hence, at (x_0, y_0), the two vectors ∇f and ∇g are parallel, i.e. scalar multiples of one another. $\square$

Example 3.1. *Suppose that you want to make a cylindrical can which has a fixed surface area of 2π units, with maximum volume. Let r be the radius of the can, and let h be its height. Then the volume of the can is $\pi r^2 h$ and the surface area is $2\pi rh + 2\pi r^2 = 2\pi r(r + h)$. Therefore what we want is to maximize the function $\pi r^2 h$ under the constraint $r(r + h) = 1$. In the language of Lagrange Multipliers, $f(r, h) = \pi r^2 h$, and $g(r, h) = r^2 + rh$. By the theorem proved above, any nondegenerate maximum satisfies the equation $\nabla f = \lambda \nabla g$ for some scalar λ. Note that $\nabla f = (2\pi rh, \pi r^2)$, and $\nabla g = (2r + h, r)$. Therefore we must solve the following three equations:*

$$2\pi rh = \lambda(2r + h) \tag{1}$$

$$\pi r^2 = \lambda r \tag{2}$$

$$r(r + h) = 1 \tag{3}$$

for the three unknowns r, h, and λ. The second equation tells me that $\lambda = \pi r$. Plugging into the first equation, $2\pi rh = 2\pi r^2 + \pi rh$, so $\pi rh = 2\pi r^2$, so $h = 2r$. Finally, plugging into the third equation, $r(3r) = 1$, so $r = \frac{1}{\sqrt{3}}$ and $h = \frac{2}{\sqrt{3}}$. Since this is the only critical point found, and also since the volume goes to 0 if r or h goes to 0, we infer that this is the maximum (although this is not a proof). We conclude that the maximum volume under the constraint is $\pi r^2 h = \frac{2\pi}{3\sqrt{3}}$.

Exercise 3.1. *Maximize the function $x^4 + y^4$ subject to the constraint $x^2 + y^2 = 1$. Are there any non-maximum critical points?*