The Implicit Function Theorem

Suppose that \( U \subset \mathbb{R}^n \) is open and \( f : U \to \mathbb{R}^m \) is a smooth map. If \( a \in U \) and \( f(a) = b \) we are interested in understanding what the inverse image \( f^{-1}(b) \) looks like near \( x = a \). The implicit function theorem has something to say about this in case \( Df(a) \) is surjective, or equivalently when the rank of \( Df(a) \) is \( m \) (one also says that \( Df(a) \) has maximal rank). Since the linear approximation to \( f(x) \) near \( x = a \) is,

\[
f(x) = f(a) + Df(a)(x - a) + O((x - a)^2),
\]

the equation \( f(x) = b = f(a) \) is approximately the same as

\[
Df(a)(x - a) = 0.
\]

To understand the implicit function theorem it will thus be useful to review the situation for solving homogeneous systems of linear equations when the number of unknowns exceeds the number of equations. An example will illustrate the points I want to make. Consider the \( 3 \times 5 \) matrix,

\[
M = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0
\end{bmatrix}.
\]

To find the null space of \( M \) we do row reduction on this matrix to put it in row-echelon form,

\[
M' = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 \\
0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0
\end{bmatrix}.
\]

The row reduced form has the same null space as \( M \) and this row-echelon form allows us to take the following attitude towards the equation for the null space \( M'x = 0 \). The pivot variables \( x_1, x_3 \) and \( x_4 \) are determined by the equations in terms of the free variables \( x_2 = a \) and \( x_5 = b \). Solving the system by back substitution we find,

\[
x_1 = -a - b,
\]

\[
x_3 = 0,
\]

\[
x_4 = 0.
\]

In this way the null space is parametrized by the free variables. Since the elements \( x \) in the null space of \( M \) represent dependence relations among the columns of \( M \) and the null space doesn’t change under the row operations that produce the row reduced form it follows that any dependence relation among the columns in the row reduced form must be reflected by the same dependence relation in the original matrix. From this one sees that a column in the original matrix will be a pivot column if and only if it is not in the span of the columns to its left in the original matrix. No reordering of the order in which the equations are written down will affect the pivot columns. However, relabeling the variables will effect the pivots. If we let \( u_1 = x_1, u_2 = x_3, \) and \( u_3 = x_4 \) with \( u_4 = x_2 \) and \( u_5 = x_5 \) then \( u_1, u_2, u_3 \) with be the pivots and \( u_4 \) and \( u_5 \) will be the free variables. This is convenient for the statement of the implicit function theorem. Suppose that \( f : U \to \mathbb{R}^m \) and for \( a \in U, Df(a) \) has maximal rank. Then by relabeling we can choose coordinates \((x, y) \in \mathbb{R}^n \) with \( x \in \mathbb{R}^m \) and \( y \in \mathbb{R}^{n-m} \) so that,

\[
Df(a) = [D_x f(a), D_y f(a)]
\]

and the \( m \times m \) matrix \( D_x f(a) \) is invertible. Here we write,

\[
D_x f = \begin{bmatrix}
\partial_1 f_1 & \cdots & \partial_m f_1 \\
\vdots & & \vdots \\
\partial_1 f_m & \cdots & \partial_m f_m
\end{bmatrix},
\]

1
and \( \partial_j = \frac{\partial}{\partial x_j} \). The implicit function theorem is,

**Theorem.** Suppose that \( U \) is open in \( \mathbb{R}^n \) and \( f : U \to \mathbb{R}^m \) is a \( C^r \) map. Suppose that \( a \in U \) and \( Df(a) \) is surjective. Order the coordinates \((x, y)\) for \( \mathbb{R}^n \) with \( x \in \mathbb{R}^m \) and \( y \in \mathbb{R}^{n-m} \) so that \( D_y f(a) \) is an \( m \times m \) invertible matrix. Write \( a = (x_0, y_0) \) for the \((x, y)\) coordinates of \( a \). Then there exists a unique \( C^r \) function \( h(y) \) defined for \( y \) in a neighborhood of \( y_0 \) so that,

\[
f(h(y), y) = f(a) = b.
\]

and \( h(y_0) = x_0 \).

Proof. Extend \( f \) to a map from \( U \) in \( \mathbb{R}^n \) by,

\[
F(x, y) = (f(x, y), y).
\]

The derivative of \( F \) at \( a \) is,

\[
DF(a) = \begin{bmatrix} D_x f(a) & D_y f(a) \\ 0 & I_{m \times m} \end{bmatrix}.
\]

This is invertible since it is easy to check that,

\[
\begin{bmatrix} A & B \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}B \\ 0 & I \end{bmatrix},
\]

and we are supposing that \( D_y f(a) \) is invertible. The inverse function theorem applies to \( F \) at \( a \) and we find there is a neighborhood of \( F(a, y_0) = (b, y_0) \) on which a \( C^r \) inverse function \( G \) is defined. Since \( F \) preserves the last \( n-m \) coordinates so does \( G \) and we find that there exist a \( C^r \) function \( h(u, v) \) with values in \( \mathbb{R}^n \) so that,

\[
G(u, v) = (h(u, v), v).
\]

Applying \( F \) to this we find,

\[
(u, v) = (f(h(u, v), v), v),
\]

or

\[
f(h(u, v), v) = u.
\]

Now fix \( u = b \) and define \( h(v) = h(b, v) \). Then,

\[
f(h(b, v), v) = b.
\]

Also since \( G(b, y_0) = a \) it follows that \( h(b, y_0) = x_0 \) so \( h(y_0) = x_0 \). Now suppose that one has another function \( \tilde{h}(y) \) which satisfies the conditions of the theorem. Then

\[
(f(h(y), y), y) = (b, y) = (f(\tilde{h}(y), y), y).
\]

Applying \( G \) to both sides we find that \( h(y) = \tilde{h}(y) \). For this to be legitimate we need to know that \((f(\tilde{h}(y), y), y)\) is in the domain of \( G \) for \( y \) in a neighborhood of \( y_0 \). But \( \tilde{h}(y_0) = x_0 \) and the continuity of \( \tilde{h} \) guarantee such a neighborhood exists. QED

Usually one cannot find an explicit form for the function \( h(y) \) but it is instructive to work out an example where this is possible to give a feeling for the result. Consider the following function,

\[
f(x, y, z) = \begin{bmatrix} x(y^2 - z^2 - 1) \\ z(x^2 - y^2 - 1) \end{bmatrix}.
\]

Suppose we are interested in the set \( Z = f^{-1}(0, 0) \). The function \( f \) was deliberately chosen so that this set is easy to determine (the components of \( f \) factor into terms with easily analysed 0's). The line \( x = 0 \) and \( z = 0 \) is in \( Z \). The lines \( z = 0 \) and \( y = \pm 1 \) are in \( Z \). The four curves \( x = \pm \sqrt{y^2 + 1} \) and \( z = \pm \sqrt{y^2 - 1} \) for
$y \geq 1$ are in $Z$. The implicit function theorem does not apply to the global description of this set but only to a local picture (or parametrization) once we’ve chosen a point $a \in \mathbb{R}^3$ at which $DF(a)$ has maximal rank. Evidently,

$$DF(x, y, z) = \begin{bmatrix} y^2 - z^2 - 1 & 2xy & -2xz \\ 2xz & 2y & x^2 - y^2 - 1 \end{bmatrix}.$$ 

Let’s look at this derivative along $x = 0$ and $z = 0$,

$$DF(0, y, 0) = \begin{bmatrix} y^2 - 1 & 0 & 0 \\ 0 & 0 & -y^2 - 1 \end{bmatrix}.$$ 

This is rank 2 at every point $y$ except $y = \pm 1$. Notice that at $y = \pm 1$ the line $y \to (0, y, 0) \in Z$ intersects with $x \to (x, \pm 1, 0) \in Z$ at $x = 0$. If $y_0 \neq \pm 1$ then the implicit function theorem tells us that near $(0, y_0, 0)$ the “pivot variables” $x$ and $z$ can be written as functions of the “free variable” $y$ in a unique way so as to keep $(x(y), y, z(y))$ in the set $Z$. Of course $x = 0$, and $z = 0$ does the trick.

Next consider,

$$DF(x, \pm 1, 0) = \begin{bmatrix} 0 & \pm 2x & 0 \\ 0 & 0 & x^2 - 2 \end{bmatrix}.$$ 

This is rank 2 except for $x = 0$ or $x = \pm \sqrt{2}$. We already know that at $x = 0$, $x \to (x, \pm 1, 0) \in Z$ collides with $y \to (0, y, 0) \in Z$ at $y = \pm 1$. At $x = \sqrt{2}$ (for example) the line $x \to (x, \pm 1, 0)$ collides with $y \to (\sqrt{y^2 + 1}, y, \sqrt{y^2 - 1}) \in Z$ for $y = \pm 1$ and $\epsilon$ can be chosen to be $\pm 1$ independently. If $x_0$ is not 0 or $\pm \sqrt{2}$ then the “pivot variables” $y$ and $z$ are locally functions of $x$ namely $y = \pm 1$ and $z = 0$ for $x$ sufficiently close to $x_0$ and $f(x, y, z) = 0$.

Finally, let’s look at one branch $x = \sqrt{y^2 + 1}$ and $z = \sqrt{y^2 - 1}$ for $y \geq 1$.

$$DF(\sqrt{y^2 + 1}, y, \sqrt{y^2 - 1}) = \begin{bmatrix} 0 & 2xy & -2xz \\ 2xz & 0 & x^2 - 2 \end{bmatrix}.$$ 

Evidently, this is rank 2 unless $z = 0$ which occurs at $y = \pm 1$. Along this curve for $y > 1$ any two of the columns of the derivative are linearly independent. Thus any pair of variables would serve as “pivots” or equivalently any variable might serve as a “free variable”. The choice of $y$ as a free variable evidently produces $x = \sqrt{y^2 + 1}$ and $z = \sqrt{y^2 - 1}$.

This example also makes it clear that life gets complicated at points $a$ where $f(a) = 0$ and $DF(a)$ is not of maximal rank. The point $(0, 1, 0)$, for example, is a sort of departure hub for different branches of the set $f^{-1}(0)$.

Our interest in the implicit function theorem is that it will provide us with lots of examples of manifolds, the arena to which we want to extend the notions of Calculus.

**Definition.** A $C^r$-manifold, $X$, is a Hausdorff space together with a collection of $C^r$ related coordinate systems. A coordinate system, also called a chart, is a homeomorphism from an open subset of $X$ to an open subset of $\mathbb{R}^n$. Every point $x \in X$ must be in the domain of some chart and any two charts $\varphi$ and $\psi$ must be related by a $C^r$ change of coordinates $\varphi \circ \psi^{-1}$.

**Theorem.** Suppose that $U$ is open in $\mathbb{R}^n$, $f : U \to \mathbb{R}^m$ is $C^r$ and $DF(a)$ is surjective for all $a \in f^{-1}(b)$. Then $f^{-1}(b)$ is a $C^r$ manifold of dimension $n - m$.

**Proof.** Since $f^{-1}(b)$ is a subset of $\mathbb{R}^n$ it is a metric space and hence also a Hausdorff space. Recall that if $a \in f^{-1}(b)$ and $DF(a)$ is surjective then for some ordering of coordinates $(x, y)$ for $\mathbb{R}^n$ (with $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^{n-m}$) about $(x_0, y_0) = a$, all the points in $f^{-1}(b)$ near $a$ are of the form $(h(y), y)$ where $h$ is a $C^r$ function defined in a neighborhood of $y_0$. We define a coordinate chart $\varphi$ in this neighborhood of $a$ in $f^{-1}(b)$ by,

$$\varphi(h(y), y) = y.$$ 

This is just the restriction to $f^{-1}(b)$ of the projection,

$$\varphi(x, y) = y,$$ 

3
from $\mathbb{R}^n$ to $\mathbb{R}^{n-m}$ and so is a continuous function. The inverse $y \rightarrow (h(y), y)$ is $C^r$ and so clearly continuous. Comparing two charts, we see that,

$$\psi \circ \varphi^{-1}(y) = \psi(h(y), y).$$

Since $\psi$ just projects onto some choice of $n-m$ coordinates this last map is clearly $C^r$ since $h$ is $C^r$. QED