The Implicit Function Theorem (IFT) is one of the most important theorems in mathematics. It is concerned with solvability of equations of the form

\[ f(x, y) = 0, \]

where \( f : X \times Y \to Y \) is some differentiable function and \( X, Y \) are some open subsets of Euclidean spaces (or even more general spaces). More precisely, it tells us when we can uniquely solve for \( y \) in terms of \( x \).

**Example.** Let \( f(x, y) = x^2 + y^2 - 1 \), where \( x, y \in \mathbb{R} \), and consider the corresponding equation \( x^2 + y^2 - 1 = 0 \).

The equation clearly defines the unit circle, \( S^1 \). When can we solve for \( y \) as a function of \( x \)? In other words, when can we express a portion of the unit circle as the graph of a function \( y = y(x) \)? (Note there’s no way we can express the whole circle as the graph of some function.) Clearly,

\[ y = \sqrt{1 - x^2} \quad \text{or} \quad y = -\sqrt{1 - x^2}. \]

The first expression is valid near any point \((x, y) \in S^1\) with positive \( y \), the second near any point with negative \( y \). However, there is no well-defined solution near \((1, 0)\) and \((-1, 0)\), simply because we can’t choose the sign of the square root. Observe that \((\pm 1, 0)\) are precisely the points at which \( \partial f/\partial y = 0 \).

To state the general, multi-variable version of the IFT we need partial derivatives of vector-valued functions. If \( X \subset \mathbb{R}^m, Y \subset \mathbb{R}^n, Z \subset \mathbb{R}^k \), and \( f : X \times Y \to Z \) is a differentiable map, then we can differentiate \( f(x, y) \) relative to \( x \) and relative to \( y \). Here’s how we do that. If \( x = (x_1, \ldots, x_m), y = (y_1, \ldots, y_n) \) and

\[ f(x, y) = (f_1(x, y), \ldots, f_k(x, y)), \]

then the partial derivatives of \( f \) with respect to \( x \) and \( y \) are the matrices

\[
D_x f = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_k}{\partial x_1} & \frac{\partial f_k}{\partial x_2} & \cdots & \frac{\partial f_k}{\partial x_m}
\end{bmatrix}, \quad D_y f = \begin{bmatrix}
\frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_n} \\
\frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_k}{\partial y_1} & \frac{\partial f_k}{\partial y_2} & \cdots & \frac{\partial f_k}{\partial y_n}
\end{bmatrix}.
\]

Observe that the \( i^{th} \) row of \( D_x f \) is the gradient of \( f \) with respect to \( x \). (Similarly for \( D_y f \).)

Note also that if \( k = n \), then \( D_y f \) is a square matrix so we can speak about its invertibility.
Theorem. Suppose \( f : X \times Y \to Y \) is differentiable and for some \((a, b) \in X \times Y\), we have 
\[
f(a, b) = 0.
\]
If \( D_y f(a, b) \) is invertible, then the equation 
\[
f(x, y) = 0
\]
has a unique solution \( y = \psi(x) \) defined in some neighborhood \( U \subset X \) of \( a \), i.e., \( f(x, \psi(x)) = 0 \), for all \( x \in U \). In particular, \( \psi(a) = b \). Furthermore, \( \psi : U \to Y \) is differentiable at every \( x \in U \) and 
\[
D\psi(x) = -[D_y f(x, \psi(x))]^{-1} D_x f(x, \psi(x)).
\]
(1)

For a proof of this theorem, see, e.g., Walter Rudin’s book *Principles of Mathematical Analysis*.

Remarks. (a) The assumption \( f(a, b) = 0 \) just means that the equation \( f(x, y) = 0 \) has a solution.

(b) If \( f \) is of class \( C^r \) (i.e., \( f \) has partial derivatives of order up to \( r \) and they are all continuous), then \( \psi \) is also \( C^1 \).

(c) Assuming that the function \( \psi \) exists and is differentiable, it is not too hard to derive (1). Here’s how that can be done. We know that 
\[
f(x, \psi(x)) = 0,
\]
for all \( x \in U \). Differentiating with respect to \( x \in U \) and using the Chain Rule, we obtain 
\[
\frac{d}{dx} f(x, \psi(x)) = D_x f(x, \psi(x)) + D_y f(x, \psi(x)) D\psi(x)
\]
\[
= 0.
\]
Rearranging the terms, we get \( D_y f(x, \psi(x)) D\psi(x) = -D_x f(x, \psi(x)) \). Since \( D_y f(x, y) \) is invertible for \((x, y)\) close to \((a, b)\), we can multiply both sides by its inverse. This yields (1).

(d) Another way of looking at IFT is this. The set of points \((x, y) \in X \times Y\) such that 
\[
f(x, y) = 0
\]
defines a **level set** \( S \) of \( f \). The conclusion of the IFT can be interpreted as saying that near any point \((x, y)\) where \( D_y f(x, y) \) is invertible, \( S \) can be represented as the **graph** of a differentiable function \( \phi \). In other words,
\[
S = \{(x, \phi(x)) : x \in U \}.
\]
Since \( \phi \) is differentiable, \( S \) is nice and smooth, i.e., it has a tangent plane at every point.

Example. Let us now look at the way IFT is applied in Hirsch-Smale-Devaney. We have a smooth flow \( \phi_t \) on \( \mathbb{R}^2 \) with the following two properties:

(i) \( \lim_{t \to 0} \phi_t(p) = 0 \) and \( \lim_{t \to -\infty} \phi_{-t}(p) = \infty \), as \( t \to \infty \), for all \( p \in \mathbb{R}^2 \setminus (0, 0) \);

(ii) the corresponding vector field \( F(p) = \frac{d}{dt} |_{t=0} \phi_t(p) \) always points inside the unit circle.
That is, at every point \( p = (\cos \theta, \sin \theta) \in S^1 \), \( F(p) \cdot N(p) < 0 \), where \( N(p) \) is the outward pointing unit normal to \( S^1 \) at \( p \in S^1 \). For \( p \in \mathbb{R}^2 \setminus (0,0) \), let \( \tau(p) \) be the time it takes the \( \phi_t \)-orbit of \( p \) to meet \( S^1 \); that is, assume

\[
|\phi_{\tau(p)}(p)| = 1.
\]

This is a well-defined function. We want to use IFT to show that \( \tau \) is differentiable. Define \( f : \mathbb{R}^2 \setminus (0,0) \times \mathbb{R} \rightarrow \mathbb{R} \) by

\[
f(p, t) = |\phi_t(p)|^2 - 1.
\]

Then \( t = \tau(p) \) is a solution of the equation \( f(p, t) = 0 \). By (i) and (ii), this solution is clearly unique. To show that \( \tau \) is differentiable, we need to verify that the \( t \)-partial of \( f \) is invertible, which in this case means nonzero. By a calculation analogous to the one in the book, we have

\[
D_t f(p, t) = 2F(\phi_t(p)) \cdot N(\phi_t(p)),
\]

which is \(< 0 \) (and therefore nonzero) when \( t = \tau(p) \). Therefore, by the IFT, \( \tau \) is differentiable.