Implicit Function Theorem for Banach Spaces

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

Theorem 1. *Let $U \subset X \times Y$ be open. Let $f\colon U \to Z$ be continuous. Assume that $D_y f(x, y)$ exists on U. Let $(x_0, y_0) \in U$ be such that $f(x_0, y_0) = 0$. Assume further that $D_y f(x_0, y_0)\colon Y \to Z$ is continuous linear isomorphism so that $D_y f(x_0 y_0)^{-1}\colon Z \to Y$ is also continuous.*

Then there exist $r, \rho > 0$ such that $B(x_0, r) \times B(y_0, \rho) \subset U$ and a function $u\colon B(x_0, r) \to B(y_0, \rho)$ such that (i) $u(x_0) = y_0$ and $f(x, u(x)) = 0$ for all $x \in B(x_0, r)$.

If we further assume that f is continuously differentiable, then u is also C^1 and we have

$$Du(x) = -D_y f(x, y)^{-1} \circ D_x f(x, u(x)), \quad \textit{for } x \in B(x_0, r).$$

Let X, Y and E be Banach spaces and let $U \subset X \times Y$ be open. The statement and the proof goes through when X, Y and E are Banach spaces. For the standard version, the reader may assume that $X := \mathbb{R}^m$ and $Y = E = \mathbb{R}^k$. Any element of $\mathbb{R}^m \times \mathbb{R}^k$ is denoted by (x, y) with $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^k$. The condition (3) of the theorem in finite dimensional case can be recast as saying that $D_y f(x_0, y_0)$ is a linear isomorphism of $\mathbb{R}^k$.

Theorem 2. *Let $f\colon U \subset X \times Y \to E$ be such that*

(1) *$f(x_0, y_0) = 0$ for some $(x_0, y_0) \in U$,*
(2) *$D_y f$, the partial derivative of f with respect to the second variable y, exists on U,*
(3) *the map $D_y f(x_0, y_0)\colon Y \to E$ is continuous and bijective*
(4) *f is continuous at (x_0, y_0).*

Then there exist r, ρ such that for every $x \in B(x_0, \rho)$, there exists a unique $y(x) \in Y$ such that $y(x) \in B(y_0, r)$ and $f(x, y(x)) = 0$.

Furthermore, if f is C^k around (x_0, y_0) so is the map $x \mapsto y(x)$ around x_0. □

Strategy: We begin with some preliminary remarks that may shed some light on the hypothesis and the proof of the theorem. To start with, we may assume without loss of generality that $(x_0, y_0) = (0, 0)$.

We first look at the case when X, Y, E are $\mathbb{R}$ and f is real analytic in U. Then we have the power series expansion of f near $(0, 0)$:

$$f(x, y) = f(0, 0) + D_x f(0, 0)x + D_y f(0, 0)y + O(2),$$

where $O(2)$ stands for higher order terms. By hypothesis, $f(x,y)=0$. Since we are looking at the set $\{(x,y) : f(x,y)=0\}$, we get $D_xf(0,0)x + D_yf(0,0)y + O(2) = 0$. This suggests to us that if we assume that $D_yf(0,0)$ is invertible, we can express y as follows:

$$y = -D_yf(0,0)^{-1}\left(D_xf(0,0)x + O(2)\right). \tag{1}$$

This tells us that we may try to adapt Newton's method of locating zeros. We recall the method briefly. Assume that x_0 is an approximate zero of $y(x)=0$ and that y has nonvanishing derivative around x_0. Then we inductively define

$$x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)}, \text{ for } n \geq 0.$$

Then x_n "converges" to a zero of y (in 'good' cases). These ideas suggest that if we define

$$y_{n+1}(x) := y_n(x) - D_yf(0,0)^{-1}f(x, y_n(x)), \text{ for } n \geq 0 \text{ where } y_0 = 0, \tag{2}$$

then y_n may converge to a solution of $f(x,y(x))=0$. This is the programme we implement by employing the contraction mapping theorem. Shedding of indices, we may rewrite (2) as follows:

$$y := y - D_yf(0,0)^{-1}f(x,y(x)), \text{ that is, } D_yf(0,0)y = D_yf(0,0)y - f(x,y). \tag{3}$$

If we define $g(x,y) := D_yf(0,0)y - f(x,y)$, then $f(x,y)=0$ iff $y = D_yf(0,0)^{-1}g(x,y)$. Thus if we define

$$T_x(y) := D_yf(0,0)^{-1}f(x,y),$$

the required solution $y(x)$ is got as a fixed point of T_x. We investigate whether the contraction mapping theorem can be applied.

Proof. We assume without loss of generality that $(x_0,y_0)=(0,0)$. Let $B := D_yf(0,0)$. By hypothesis, B is a continuous bijection of Y onto E.

Let $g(x,y) := By - f(x,y)$. Note that $D_yg(x,y) = D_yf(0,0) - D_yf(x,y)$, so that $D_yg(0,0)=0$. Since f and D_yf are continuous at 0, we can apply mean value inequality to g. By the mean value inequality, we have

$$\begin{aligned} \| g(x,y) - g(x,z) \| &\leq \sup_{0\leq t\leq 1} \| D_yg(x, z+t(y-z)) \| \, \| y-z \| \\ &\leq \sup_{0\leq t\leq 1} \| D_yf(0,0) - D_yf(x, z+t(y-z)) \| \, \| y-z \| . \end{aligned} \tag{4}$$

If we now assume that $\| x \|, \| y \|, \| z \| \leq r$, then (4) implies that

$$\| g(x,y) - g(x,z) \| = o(1), \text{ as } r \to 0, \tag{5}$$

by the continuity of D_yf at 0.

Let $T_x(y) := B^{-1}g(x,y)$. We have

$$\| T_xy - T_xz \| \leq o(1) \left\| D_yf(0,0)^{-1} \right\| \| y-z \| \leq 2^{-1} \| y-z \| . \tag{6}$$

Thus, the family $\{T_x\}$ is a family of *uniformly contraction maps*, (uniformly in x) of the complete metric space $B[0, r]$ to itself. The contraction mapping theorem now yields the existence and the uniqueness of the function y.

We now establish the continuity of y as follows, where we shall write $T(x, y)$ for $T_x y$ etc.

$$\begin{aligned} \| y(x+h) - y(x) \| &= \| T(x+h, y+h) - T(x, y \| \\ &= \| T(x+h, y+h) - T(x+h, y) \| + \| T(x+h, y) - T(x, y) \| \\ &\leq 2^{-1} \| y(x+h) - y(x) \| + \| T(x+h, y) - T(x, y) \| , \text{ by (6).} \end{aligned}$$

Transferring the first term on the right to the left side, we obtain

$$\| y(x+h) - y(x) \| \leq 2 \| T(x+h, y(x)) - T(x, y(x)) \| . \tag{7}$$

We now show that the map $x \mapsto y(x)$ is Lipschitz. We start with (7).

$$\begin{aligned} \| y(x+h) - y(x) \| &\leq 2 \| T(x+h, y(x)) - T(x, y(x)) \| \\ &\leq 2 \left\| D_y f(0,0)^{-1} \left[f(x+h, y) - f(x, y) \right] \right\| \\ &\leq \left\| D_y f(0,0)^{-1} \right\| \sup_x \| Df(x, y) \| \| h \| . \end{aligned} \tag{8}$$

We now show that $x \mapsto y(x)$ is differentiable.

$$\begin{aligned} 0 &= f(x+th, y(x+th)) - f(x, y(x)) \text{ near } (0,0) \\ &= D_x f(x, y(x)) th + D_y f(x, y(x)) \left[y(x+th) - y(x) \right] + o(t), \end{aligned}$$

as $t \to 0$ by the Lipschitz continuity of y. Hence

$$y(x+th) - y(x) = -t D_y f(x, y(x))^{-1} \cdot D_x f(x, y(x)) h + o(t),$$

as $t \to 0$. Here we have used the continuity of y, Df at $(0,0)$ to conclude that $D_y f(x, y(x))$ is invertible near (0,0). Thus the directional derivatives $D_h y$ exist for $h \in X$ and they are given by

$$D_h y(x) = -D_y f(x, y(x))^{-1} D_x f(x, y(x)) h.$$

The continuity of $Df(x, y)$ implies that of $D_x f(x, y)$ and $D_y f(x, y)$ around $(0,0)$. We therefore conclude that y is C^1 with

$$Dy(x) = -D_y f(x, y(x))^{-1} \circ D_x f(x, y(x)).$$

□