INROADS

Real Analysis Exchange
Vol. 39(1), 2013/2014, pp. 207–218

Oswaldo de Oliveira, Department of Mathematics, Universidade de São Paulo, São Paulo, SP 05508-090, Brasil. email: oliveira@ime.usp.br

THE IMPLICIT AND THE INVERSE FUNCTION THEOREMS: EASY PROOFS

Abstract

This article presents simple and easy proofs of the Implicit Function Theorem and the Inverse Function Theorem, in this order, both of them on a finite-dimensional Euclidean space, that employ only the Intermediate-Value Theorem and the Mean-Value Theorem. These proofs avoid compactness arguments, the contraction principle, and fixed-point theorems.

1 Introduction.

The objective of this paper is to present very simple and easy proofs of the Implicit and Inverse Function theorems, in this order, on a finite-dimensional Euclidean space. The lack of sophisticated tools used in its proof could make The Implicit Function Theorem more acessible to an undergraduate audience. Besides following Dini's inductive approach, these demonstrations do not employ compactness arguments, the contraction principle or any fixed-point theorem. Instead of such tools, these proofs rely on the Intermediate-Value Theorem and the Mean-Value Theorem on the real line.

The history of the Implicit and Inverse Function theorems is quite long and dates back to R. Descartes (on algebraic geometry), I. Newton, G. Leibniz, J. Bernoulli, and L. Euler (and their works on infinitesimal analysis), J. L. Lagrange, A. L. Cauchy, and U. Dini (on functions of real variables and differential geometry). Let us discuss briefly some of the techniques that have been used to prove these theorems.

Mathematical Reviews subject classification: Primary: 26B10, 26B12; Secondary: 97140
Key words: Implicit Function Theorems, Calculus of Vector Functions, Differential Calculus, Functions of Several Variables
Received by the editors February 28, 2013
Communicated by: Emma D'Aniello

Newton's iteration method for inverting $f : \mathbb{R} \to \mathbb{R}$ near x_0, with $f'(x_0) \neq 0$, shows a sequence converging to a solution x of the equation $y = f(x)$. Through the linearization $y \approx f(x_0)+f'(x_0)(x-x_0)$ we obtain the approximate solution $x \approx x_0 + f'(x_0)^{-1}[y - f(x_0)]$. Then, we use the Newton-Raphson iteration $x_{n+1} = x_n + f'(x_n)^{-1}[y - f(x_n)]$, with $n = 0, 1, 2, \ldots$. See Dontchev and Rockafellar [4, pp. 11–14] for a proof of this method in $\mathbb{R}^n$ that employs compactness.

Lagrange's inversion formula shows the formal Taylor series of the local inverse of an analytic function $f(z)$ such that $f'(z_0) \neq 0$. On the other hand, Cauchy's proof of the Implicit Function Theorem (for complex functions) is considered the first rigorous proof of this theorem. By employing the method of residues, Cauchy gave an integral representation for the solution. He also proved such theorem by the method of the majorants (a technique used to proof the Cauchy-Kowalewski theorem for analytic partial differential equations), which also applies to real analytic functions. See Burckel [2, pp. 173–174, 180–183] and Krantz and Parks [6, pp. 30–38] and [7, pp. 27–32, 117–121].

The two most usual approaches to the Implicit and Inverse Function theorems on a finite-dimensional Euclidean space begin with a proof of the latter (then, the former follows). Hence, let us consider a function $F : \mathbb{R}^n \to \mathbb{R}^n$ of class C^1 and a point x_0 such that the differential $DF(x_0)$ is invertible.

The most basic of these techniques uses elementary calculus and holds only in finite dimensions, since it employs the local compactness of $\mathbb{R}^n$. Let us outline a proof. There is $m > 0$ such that $\|DF(x_0)(v)\| \geq m|v|$, for all v in $\mathbb{R}^n$. Hence, $\|DF(x) - DF(x_0)\| \leq m(2\sqrt{n})^{-1}$, with $DF(x)$ invertible, for all x in an open ball $B = B(x_0; r)$, $r > 0$. By applying the mean-value theorem to each component of F we find $|F(x_1) - F(x_2)| \geq m|x_1 - x_2|/2$, for all x_1, x_2 in the closure $\overline{B}$ of B. Thus, $F : \overline{B} \to F(\overline{B})$ is bicontinuous. If ∂B is the boundary of B, then the distance d of $F(x_0)$ to the compact $F(\partial B)$ is positive. Given y' in the open ball $V = B(F(x_0); d/2)$, we put $\varphi(x) = |y' - F(x)|^2$, for all x in $\overline{B}$. We have $|y' - F(x)| > |y' - F(x_0)|$, for all x in ∂B, and through Weierstrass's Theorem on Minima we see that φ has a minimum at a x' in B. By differentiating φ, we prove $F(x') = y'$. Thus, $U = B \cap F^{-1}(V)$ is open and $F : U \to V$ has a continuous inverse G. Given y and $y+k$, both in V, we write $G(y) = x$ and $G(y + k) = x + h$. Hence, $k \to 0$ if and only if $h \to 0$. Putting $S = DF(x)$, there is a $c > 0$ satisfying $|S(h/|h|)| \geq c$ for all $h \neq 0$. We also have $k = F(x+h) - F(x) = S(h) + |h|E(h)$, where $E(h) \to 0$ as $h \to 0$. Thus,

$$\lim_{k\to 0} \frac{G(y + k) - G(y) - S^{-1}(k)}{|k|} = \lim_{h\to 0} \frac{-S^{-1}(E(h))}{|S(h/|h|) + E(h)|} = 0.$$

That is, G is differentiable at y and $DG(y) = DF(x)^{-1}$, where $x = G(y)$. For details, see Knapp [5, pp. 152–161] and Spivak [10, pp. 40–45].

The second approach is more advanced and abstract, relies on the completeness of $\mathbb{R}^n$, can be extended to complete normed spaces (Banach spaces) of arbitrary dimension, and holds in more general spaces than C^k functions. The technique depends on basic functional analysis and resembles Newton's method since it employs a somewhat similar iterative procedure, which we now state.

The Contraction Mapping Principle. *Let X be a complete metric space, with metric d. Let us suppose that $\Phi : X \to X$ satisfies $d(\Phi(x), \Phi(y)) \leq \lambda d(x, y)$, for all x, y in X, where λ is a constant and $0 < \lambda < 1$. Then, Φ has a unique fixed point. That is, there exists a unique x in X such that $\Phi(x) = x$.*

Let us summarize a proof of the Inverse Function Theorem that employs this principle. Searching for a solution x of $F(x) = y$, near x_0, we define $\Phi(x) = x + T^{-1}[y - F(x)]$, with $T = DF(x_0)$ and y a parameter. Hence, $F(x) = y$ is equivalent to $\Phi(x) = x$. Since $D\Phi(x_0) = 0$, we have $\|D\Phi(x)\| < 2^{-1}$ for all x in an open ball U containing x_0. We may assume that $DF(x)$ is invertible at every x in U. The mean-value inequality yields $|\Phi(x_1) - \Phi(x_2)| \leq 2^{-1}|x_1 - x_2|$, for all x_1, x_2 in U. Therefore, Φ has at most one fixed point in U and thus F is injective on U. Hence, $F : U \to F(U)$ has an inverse G. Let us see that $V = F(U)$ is open. Given $y_3 = F(x_3)$, with x_3 in U, we pick an open ball $B = B(x_3; r)$, with $r > 0$, whose closure $\overline{B}$ lies in U. Fixing y such that $|y - y_3| < 2^{-1}\|T^{-1}\|^{-1}r$ and taking any x in $\overline{B}$, we have $|\Phi(x) - \Phi(x_3)| \leq r/2$, $|\Phi(x_3) - x_3| \leq \|T^{-1}\|2^{-1}\|T^{-1}\|^{-1}r = r/2$, and $|\Phi(x) - x_3| \leq r$. Thus, Φ is a contraction of the complete set $\overline{B}$ into $\overline{B}$. Hence, Φ has a fixed point x in $\overline{B}$ and $F(x) = y$. Therefore, V is open. Analogously, given any open subset of U, its image by F is an open subset of V. Thus, $F : U \to V$ is bicontinuous. We proved above that G is differentiable. For details, see Rudin [8, pp. 221–228]. A proof of the Implicit Function Theorem in Banach spaces, based on the contraction mapping principle, is given by Krantz and Parks [7, pp. 48–52].

The implicit and inverse function theorems are also true on manifolds and other settings. Moreover, they hold in many classes of functions (e.g., C^k, $C^{k,\alpha}$, Lipschitz, analytic). For extensive accounts on the history of the Implicit Function Theorem and further developments (as in differentiable manifolds, Riemannian geometry, partial differential equations, etc.), see Krantz and Parks [7] (this book includes a proof of a version of the powerful Nash-Moser theorem), Dontchev and Rockafellar [4, pp. 7–8, 57–59], and Scarpello [9].

In this article, we prove by induction the Implicit Function Theorem and from it we obtain the Inverse Function Theorem. This approach is accredited to U. Dini (1876), who was the first to present a proof of the Implicit Function Theorem for a system with several equations and several real variables, and then stated and proved the Inverse Function Theorem. See Dini [3, pp. 197–241].

Another proof by induction of the Implicit Function Theorem, that also simplifies Dini's argument, is given by Krantz and Parks [7, pp. 36–41]. However, this particular proof by Krantz and Parks does not establish the local uniqueness of the implicit solution of the given equation. On the other hand, the proof presented in this paper further simplifies Dini's argument and makes the whole proof of the Implicit Function Theorem very simple, easy, and with very few computations. The Inverse Function Theorem then follows immediately.

2 Notations and Preliminaries.

We assume without proof the following basic theorems.

The Intermediate-Value Theorem. *Let $f : [a, b] \to \mathbb{R}$ be continuous. If λ is a value between $f(a)$ and $f(b)$, then there is a c in $[a, b]$ satisfying $f(c) = \lambda$.*

The Mean-Value Theorem. *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on the open interval (a, b). Then, there exists c in (a, b) satisfying $f(b) - f(a) = f'(c)(b - a)$.*

Let us consider n and m, both in $\mathbb{N}$. In what follows we fix the ordered canonical bases $\{e_1, \ldots, e_n\}$ and $\{f_1, \ldots, f_m\}$, of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Given $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, both in $\mathbb{R}^n$, their inner product is $\langle x, y\rangle = x_1y_1 + \cdots + x_ny_n$. The norm of x is $|x| = \sqrt{\langle x, x\rangle}$ and the open ball centered at x and radius $r > 0$ is $B(x; r) = \{y \text{ in } \mathbb{R}^n : \ |y - x| < r\}$.

We identify a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ with the $m \times n$ matrix $M = (a_{ij})$, where $T(e_j) = a_{1j}f_1 + \cdots + a_{mj}f_m$, for each $j = 1, \ldots, n$. The norm of T is $\|T\| = \sup\{|T(v)| : |v| \leq 1\}$ and we have $|T(v)| \leq \|T\||v|$, for all v in $\mathbb{R}^n$. Hence, T is continuous everywhere. We also write Tv for $T(v)$.

Let Ω be an open set in $\mathbb{R}^n$. Given a function $F : \Omega \to \mathbb{R}^m$, we denote by $F_i : \Omega \to \mathbb{R}$ the ith component of F, for each $i = 1, \ldots, m$. We say that F is differentiable at p in Ω if there is a linear map $DF(p) : \mathbb{R}^n \to \mathbb{R}^m$ and a function $E : B(0; r) \to \mathbb{R}^m$ defined on some $B(0; r)$, with $r > 0$, such that $F(p + h) = F(p) + DF(p)(h) + E(h)|h|$, for all $|h| < r$, where $E(h) \to 0$ as $h \to 0$ and $E(0) = 0$. The function F is differentiable if it is differentiable at all points in Ω. The matrix identified with $DF(p)$ is the Jacobian matrix of

F at p,

$$JF(p) = \left(\frac{\partial F_i}{\partial x_j}(p)\right)_{\substack{1\le i\le m\\ 1\le j\le n}} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_n}(p) \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(p) & \cdots & \frac{\partial F_m}{\partial x_n}(p) \end{pmatrix}$$

[if $m = 1$, then we write $JF(p) = \nabla F(p)$]. We say that F is of class C^1 if F and its first-order partial derivatives are continuous on Ω. In such case, we also say that F is in $C^1(\Omega;\mathbb{R}^m)$.

The following lemma (a particular case of the chain rule but sufficient for our purposes) is a local result. For practicality, we enunciate it for a function F defined on $\mathbb{R}^n$.

Lemma 1. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable, $T : \mathbb{R}^k \to \mathbb{R}^n$ be the linear function associated to a $n \times k$ real matrix M, and y be a fixed point in $\mathbb{R}^n$. Then, the function $G(x) = F(y + Tx)$, where x is in $\mathbb{R}^k$, is differentiable and satisfies $JG(x) = JF(y + Tx)M$, for all x in $\mathbb{R}^k$.*

PROOF. Let us fix x in $\mathbb{R}^k$. Given v in $\mathbb{R}^n$, by the differentiability of F we have $F(y+Tx+v) = F(y+Tx)+DF(y+Tx)v+E(v)|v|$, where $E(v) \to 0$ as $v \to 0$. Substituting $v = Th$, where h is in $\mathbb{R}^k$, into the last identity we obtain $G(x + h) = G(x) + DF(y + Tx)Th + E(Th)|Th|$. Thus, supposing $h \neq 0$, we have $|\frac{E(Th)|Th|}{|h|}| \leq \frac{|E(Th)|\|T\||h|}{|h|} = \|T\||E(Th)|$. If $h \to 0$, then $Th \to 0$ and $E(Th) \to 0$. Hence, G is differentiable at x and $JG(x) = JF(y + Tx)M$. $\square$

With the hypothesis on Lemma 1, we see that if F is C^1 then G is also C^1.

Given a and b, both in $\mathbb{R}^n$, we put $\overline{ab} = \{a + t(b - a) : 0 \leq t \leq 1\}$. The following lemma, the mean-value theorem in several variables, is a trivial consequence of the mean-value theorem on the real line and thus we omit the proof.

Lemma 2. *Let us consider $F : \Omega \to \mathbb{R}$ differentiable, with Ω open in $\mathbb{R}^n$. Let a and b be points in Ω such that the segment $\overline{ab}$ is within Ω. Then, there exists c in $\overline{ab}$ satisfying*

$$F(b) - F(a) = \langle \nabla F(c), b - a\rangle .$$

We denote the determinant of a real square matrix M by $\det M$.

Lemma 3. *Let F be in $C^1(\Omega;\mathbb{R}^n)$, with Ω open within $\mathbb{R}^n$, and p in Ω satisfying $\det JF(p) \neq 0$. Then, F restricted to some ball $B(p;r)$, with $r > 0$, is injective.*

PROOF. (See Bliss [1]) Since F is of class C^1 and the determinant function $\det : \mathbb{R}^{n^2} \to \mathbb{R}$ is continuous and $\det JF(p) = \det\big(\frac{\partial F_i}{\partial x_j}(p)\big) \neq 0$, there is $r > 0$ such that $\det\big(\frac{\partial F_i}{\partial x_j}(\xi_{ij})\big)$ does not vanish, for all ξ_{ij} in $B(p;r)$, where $1 \leq i, j \leq n$.

Let a and b be distinct in $B(p;r)$. By employing the mean-value theorem in several variables to each component F_i of F, we find c_i in the segment $\overline{ab}$, within $B(p;r)$, such that $F_i(b) - F_i(a) = \langle \nabla F_i(c_i), b - a\rangle$. Hence,

$$\begin{pmatrix} F_1(b) - F_1(a) \\ \vdots \\ F_n(b) - F_n(a) \end{pmatrix} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(c_1) & \cdots & \frac{\partial F_1}{\partial x_n}(c_1) \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1}(c_n) & \cdots & \frac{\partial F_n}{\partial x_n}(c_n) \end{pmatrix} \begin{pmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{pmatrix}.$$

Since $\det\big(\frac{\partial F_i}{\partial x_j}(c_i)\big) \neq 0$ and $b - a \neq 0$, we conclude that $F(b) \neq F(a)$. □

3 The Implicit and the Inverse Function Theorems.

The first implicit function result we prove concerns one equation and several variables. We denote the variable in $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$ by (x, y), where $x = (x_1, \ldots, x_n)$ is in $\mathbb{R}^n$ and y is in $\mathbb{R}$.

Theorem 4. *Let $F : \Omega \to \mathbb{R}$ be of class C^1 in an open set Ω inside $\mathbb{R}^n \times \mathbb{R}$ and (a, b) be a point in Ω such that $F(a, b) = 0$ and $\frac{\partial F}{\partial y}(a, b) > 0$. Then, there exist open sets $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}$, with $(a, b) \in X \times Y \subset \Omega$, satisfying the following.*

- *There is a unique $f : X \to Y$ such that $F\big(x, f(x)\big) = 0$, for all $x \in X$.*
- *We have $f(a) = b$. Moreover, the function f is of class C^1 and satisfies*

$$\frac{\partial f}{\partial x_j}(x) = -\frac{\frac{\partial F}{\partial x_j}(x, f(x))}{\frac{\partial F}{\partial y}(x, f(x))}, \text{ for all } x \text{ in } X, \text{ where } j = 1, \ldots, n.$$

PROOF. Let us split the proof into three parts: existence and uniqueness, continuity, and differentiability.

◇ Existence and Uniqueness. Since $\frac{\partial F}{\partial y}(a, b) > 0$, by continuity there exists a non-degenerate $(n+1)$-dimensional parallelepiped $X' \times [b_1, b_2]$, centered at (a, b) and contained in Ω, whose edges are parallel to the coordinate axes such that $\frac{\partial F}{\partial y} > 0$ on $X' \times [b_1, b_2]$. Then, the function $F(a, y)$, where y runs over $[b_1, b_2]$, is strictly increasing and $F(a, b) = 0$. Thus, we have $F(a,b_1) < 0$ and $F(a,b_2) > 0$. By the continuity of F, there exists an open non-degenerate n-dimensional parallelepiped X, centered at a and contained in X', whose edges are parallel to the coordinate axes such that for every x in X we have $F(x,b_1) < 0$ and $F(x,b_2) > 0$. Hence, fixing an arbitrary x in X and employing the intermediate-value theorem on the strictly increasing function $F(x,y)$, where y runs over $[b_1,b_2]$, yields the existence of a unique $y = f(x)$ inside the open interval $Y = (b_1,b_2)$ such that $F(x,f(x)) = 0$.

- ◇ Continuity. Let $\overline{b_1}$ and $\overline{b_2}$ be such that $b_1 < \overline{b_1} < b < \overline{b_2} < b_2$. From above, there exists an open set X'', contained in X and containing a, such that $f(x)$ is in the open interval $(\overline{b_1},\overline{b_2})$, for all x in X''. Thus, f is continuous at $x = a$. Now, given any a' in X, we put $b' = f(a')$. Then, $f : X \to Y$ is a solution of the problem $F(x,h(x)) = 0$, for all x in X, with the condition $h(a') = b'$. Thus, from what we have just done it follows that f is continuous at a'.

- ◇ Differentiability. [At this point in the proof, Dini went on to use a compactness argument, whereas we will use the mean-value theorem instead.] Given x in X and j in $\{1,\ldots,n\}$, let e_j be the jth canonical vector in $\mathbb{R}^n$ and $t \neq 0$ be small enough so that $x+te_j$ is in X. Putting $P = \big(x,f(x)\big)$ and $Q = \big(x+te_j, f(x+te_j)\big)$, we have $F(P) = 0 = F(Q)$. Moreover, $Q-P = \big(0,\ldots,0,t,0,\ldots,0,f(x+te_j)-f(x)\big)$ is in $\mathbb{R}^{n+1}$, where t is the jth coordinate of $Q-P$. Thus, by employing the mean-value theorem in several variables on F restricted to the segment $\overline{PQ}$ within the open set $X \times Y$, we find a point $(\overline{x},\overline{y})$, depending on t and inside $\overline{PQ}$, satisfying

$$\begin{aligned} 0 &= F(Q) - F(P) = \langle \nabla F(\overline{x},\overline{y}), Q-P \rangle \\ &= \tfrac{\partial F}{\partial x_j}(\overline{x},\overline{y})t + \tfrac{\partial F}{\partial y}(\overline{x},\overline{y})[f(x+te_j) - f(x)]. \end{aligned}$$

Since f is continuous, we have $(\overline{x},\overline{y}) \to P = (x,f(x))$ as $t \to 0$. Moreover, $\frac{\partial F}{\partial x_j}$ and $\frac{\partial F}{\partial y}$ are continuous, with $\frac{\partial F}{\partial y}$ not vanishing on $X \times Y$. Thus, by employing the identities displayed right above we conclude that

$$\lim_{t\to 0} \frac{f(x+te_j)-f(x)}{t} = \lim_{t\to 0} -\frac{\frac{\partial F}{\partial x_j}(\overline{x},\overline{y})}{\frac{\partial F}{\partial y}(\overline{x},\overline{y})} = -\frac{\frac{\partial F}{\partial x_j}(x,f(x))}{\frac{\partial F}{\partial y}(x,f(x))}.$$

This gives the desired formula for $\frac{\partial f}{\partial x_j}$ and implies that f is of class C^1.

$\square$

Next, we prove the general implicit function theorem. In general, we apply this theorem when we have a nonlinear system with m equations and $n+m$ variables. Analogously to a linear system, we interpret n variables as independent variables and determine the remaining m variables, called dependent variables, as a function of the n independent variables.

Let us introduce some helpful notation. As before, we denote by $x = (x_1, ..., x_n)$ a point in $\mathbb{R}^n$ and by $y = (y_1, \ldots, y_m)$ a point in $\mathbb{R}^m$. Given Ω an open subset of $\mathbb{R}^n \times \mathbb{R}^m$ and a differentiable function $F : \Omega \to \mathbb{R}^m$, we write $F = (F_1, \ldots, F_m)$, with F_i the ith component of F and $i = 1, \ldots, m$. We put

$$\frac{\partial F}{\partial y} = \left(\frac{\partial F_i}{\partial y_j}\right)_{\substack{1\le i\le m\\ 1\le j\le m}} = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}.$$

Analogously, we define the matrix $\frac{\partial F}{\partial x} = \left(\frac{\partial F_i}{\partial x_k}\right)$, with $1 \le i \le m$ and $1 \le k \le n$.

Theorem 5. (The Implicit Function Theorem). *Let F be in $C^1(\Omega; \mathbb{R}^m)$, with Ω an open set in $\mathbb{R}^n \times \mathbb{R}^m$, and (a, b) a point in Ω such that $F(a, b) = 0$ and $\frac{\partial F}{\partial y}(a, b)$ is invertible. Then, there exist an open set X, inside $\mathbb{R}^n$ and containing a, and an open set Y, inside $\mathbb{R}^m$ and containing b, satisfying the following.*

- *Given x in X, there is a unique $y = f(x)$ in Y such that $F(x, f(x)) = 0$.*
- *We have $f(a) = b$. Moreover, $f : X \to Y$ is of class C^1 and*

$$Jf(x) = -\left[\frac{\partial F}{\partial y}(x, f(x))\right]^{-1}_{m\times m} \left[\frac{\partial F}{\partial x}(x, f(x))\right]_{m\times n}, \text{ for all } x \text{ in } X.$$

Proof. We split the proof into four parts: finding Y, existence and differentiability, differentiation formula, and uniqueness.

$\diamond$ Finding Y. Defining $\Phi(x, y) = \big(x, F(x, y)\big)$, where (x, y) is in Ω, we have

$$\det J\Phi = \det \left(\begin{array}{c|c} I & 0 \\ \hline \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \end{array}\right) = \det \frac{\partial F}{\partial y},$$

with I the identity matrix of order n and 0 the $n \times m$ zero matrix. Thus, $\det J\Phi(a, b) \neq 0$. As a consequence, shrinking Ω if needed, by Lemma 3 we may assume that Φ is injective and $\Omega = X' \times Y$, with X' an open set in $\mathbb{R}^n$ that contains a and Y an open set in $\mathbb{R}^m$ that contains b.

◇ Existence and differentiability. We claim that the equation $F(x, h(x)) = 0$, with the condition $h(a) = b$, has a solution $f = f(x)$ of class C^1 on some open set containing a. Let us prove it by induction on m.

The case $m = 1$ follows from Theorem 4. Let us assume that the claim is true for $m - 1$. Then, for the case m, following the notation $F = (F_1, \ldots, F_m)$ we write $\mathcal{F} = (F_2, \ldots, F_m)$. Furthermore, we put $(x; y) = (x_1, \ldots, x_n; y_1, \ldots, y_m)$, $y' = (y_2, \ldots, y_m)$, $y = (y_1; y')$, and $(x; y) = (x; y_1; y')$.

Let us consider the invertible matrix $J = \frac{\partial F}{\partial y}(a, b)$ and the associated bijective linear function $\mathcal{J} : \mathbb{R}^m \to \mathbb{R}^m$. By Lemma 1 we deduce that the function $G(x; z) = F[x; b + \mathcal{J}^{-1}(z - b)]$, defined in some open subset of $\mathbb{R}^n \times \mathbb{R}^m$ that contains (a, b), satisfies (interpreting the variable x as a fixed parameter) the identity $\frac{\partial G}{\partial z}(x; z) = \frac{\partial F}{\partial y}[x; b + \mathcal{J}^{-1}(z - b)]J^{-1}$. Hence, the function G satisfies $\frac{\partial G}{\partial z}(a; b) = JJ^{-1}$ and the condition $G(a; b) = 0$. Therefore, we may assume that J is the identity matrix of order m.

Now, let us consider the equation $F_1(x; y_1; y') = 0$, where x and y' are independent variables and y_1 is a dependent variable, with the condition $y_1(a; b') = b_1$. Since $\frac{\partial F_1}{\partial y_1}(a; b_1; b') = 1$, there exists by Theorem 4 a function $\varphi(x; y')$ of class C^1 on some open set [let us say, a cartesian product $U \times V$ of open sets] containing $(a; b')$ that satisfies

$$F_1[x; \varphi(x; y'); y'] = 0 \text{ and the condition } \varphi(a; b') = b_1,$$

on this open set. Next, substituting $y_1 = \varphi(x; y')$ into $\mathcal{F}(x; y_1; y') = 0$, we look at solving the equation

$$\mathcal{F}[x; \varphi(x, y'); y'] = 0, \text{ with the condition } y'(a) = b'.$$

Differentiating $\mathcal{F}[x; \varphi(x; y'); y']$, with respect to $y_2, \ldots, y_m$, we find

$$\frac{\partial F_i}{\partial y_1}(a; b)\frac{\partial \varphi}{\partial y_j}(a; b') + \frac{\partial F_i}{\partial y_j}(a; b) = \ 0 + \frac{\partial F_i}{\partial y_j}(a; b), \text{ where } 2 \leq i, j \leq m.$$

The matrix $\left(\frac{\partial F_i}{\partial y_j}(a; b)\right)$, where $2 \leq i, j \leq m$, is the identity matrix of order $m - 1$. Hence, by induction hypothesis there is a function ψ of class C^1 on an open set X containing a [with the image of ψ inside V] that satisfies

$$\mathcal{F}\big[x; \varphi\big(x; \psi(x)\big), \psi(x)\big] = 0, \text{ for all } x \text{ in } X, \text{ and the condition } \psi(a) = b'.$$

We also have $F_1\big[x;\varphi\big(x;\psi(x)\big);\psi(x)\big]=0$, for all x in X. Defining $f(x)=\big(\varphi(x;\psi(x));\psi(x)\big)$, with x in X, we obtain $F[x;f(x)]=0$, for all x in X, and $f(a)=\big(\varphi(a;b');b'\big)=(b_1;b')=b$, where f is of class C^1 on X.

- Differentiation formula. Differentiating $F[x,f(x)]=0$ we find

$$\frac{\partial F_i}{\partial x_k}+\sum_{j=1}^{m}\frac{\partial F_i}{\partial y_j}\frac{\partial f_j}{\partial x_k}=0,\ \text{with } 1\le i\le m \text{ and } 1\le k\le n.$$

  In matrix form, we write $\frac{\partial F}{\partial x}\big(x,f(x)\big)+\frac{\partial F}{\partial y}\big(x,f(x)\big)Jf(x)=0$.

- Uniqueness. Let $g:X\to Y$ be a function satisfying $F\big(x,g(x)\big)=0$, for all x in X, and $g(a)=b$. Given an arbitrary x in X, following the definition of Φ we have $\Phi(x,g(x))=\big(x,F(x,g(x)\big)=(x,0)$ and $\Phi\big(x,f(x)\big)=\big(x,F(x,f(x)\big)=(x,0)$. Since in the first part of this proof (the "finding Y" part) we established that Φ is injective, we deduce the identity $\big(x,g(x)\big)=\big(x,f(x)\big)$, for all x in X. Thus, $g=f$.

$\square$

Theorem 6. (The Inverse Function Theorem). *Let $F:\Omega\to\mathbb{R}^n$, where Ω is an open set in $\mathbb{R}^n$, be of class C^1 and p a point in Ω such that $JF(p)$ is invertible. Then, there exist an open set X containing p, an open set Y containing $F(p)$, and a function $G:Y\to X$ of class C^1 that satisfies $F\big(G(y)\big)=y$, for all y in Y, and $G\big(F(x)\big)=x$, for all x in X. Moreover,*

$$JG(y)=JF\big(G(y)\big)^{-1},\ \textit{for all } y \textit{ in } Y.$$

Proof. We split the proof into two parts: existence and differentiation formula.

- Existence. Shrinking Ω, if necessary, by Lemma 3 we may assume that F is injective. The function $\Phi(x,y)=F(x)-y$, where (x,y) is in $\Omega\times\mathbb{R}^n$, is of class C^1 and satisfies $\Phi\big(p,F(p)\big)=0$ and $\frac{\partial\Phi}{\partial x}\big(p,F(p)\big)=JF(p)$. From the implicit function theorem it follows that there exist an open set Y containing $F(p)$ and a function $G:Y\to\Omega$ of class C^1 such that $\Phi\big(G(y),y\big)=F\big(G(y)\big)-y=0$, for all y in Y. That is, we have $F\big(G(y)\big)=y$, for all y in Y.

  Hence, the set Y is contained in the image of F. Since F is continuous and injective, the pre-image set $X=F^{-1}(Y)$ is open, contains p, and F maps X bijectively onto Y.

  The identity $F\big(G(y)\big)=y$, for all y in Y, implies that G maps Y to X. Since F is bijective from X to Y, the map G is bijective from Y to X.

◇ Differentiation formula. Let us write $F(x) = \big(F_1(x), \dots, F_n(x)\big)$ and $G(y) = \big(G_1(y), \dots, G_n(y)\big)$. Differentiating $\big(G_1(F(x)), \dots, G_n(F(x))\big)$ we obtain

$$\sum_{k=1}^{n} \frac{\partial G_i}{\partial y_k} \frac{\partial F_k}{\partial x_j} = \frac{\partial x_i}{\partial x_j} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

□

4 Some Final Remarks.

It is the author's belief that, in general, an elementary and easy proof of a theorem can help very much a beginner in getting a good understanding of such theorem. Thus, the author hopes that this proof of the Implicit Function Theorem provides a good alternative introduction to this fundamental theorem of Real Analysis.

Acknowledgment. I thank Professors R. B. Burckel, A. Lymberopoulos , P. A. Martin, and J. V. Ralston for their very valuable comments and suggestions. I also thank Professor R. Bianconi, for translating part of a Dini's book, and the referee for several constructive suggestions.

References

[1] G. A. Bliss, *A new proof of the existence theorem for implicit functions*, Bull. Amer. Math. Soc., **18(4)** (1912), 175–179.

[2] R. B. Burckel, *An Introduction to Classical Complex Analysis*, **1**, Birkäuser Verlag, Basel, Stuttgart, 1979.

[3] U. Dini, *Lezione di Analisi Infinitesimale*, **1**, Pisa, 1907. (Italian).

[4] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer, New York, 2009.

[5] A. W. Knapp, *Basic Real Analysis*, Birkhäuser, Boston, 2005.

[6] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, Birkhäuser Verlag, Basel, 1992.

[7] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem - History, Theory, and Applications*, Birkhäuser, Boston, 2002.

[8] W. Rudin, *Principles of Mathematical Analysis*, Third edition, McGraw-Hill, Beijing, China, 2003.

[9] *G. M. Scarpello, A historical outline of the theorem of implicit functions*, Divulg. Mat., **10(2)** (2002), 171–180.

[10] M. Spivak, *Calculus on Manifolds*, Perseus Books, Cambridge, Massachusetts, 1965.