AJMAS
https://journal.utripoli.edu.ly/index.php/Alqalam/index eISSN 2707-7179

Original article

Exploring the Connections Between the Implicit Function Theorem and Optimization Methods

Idris Abdulhamid[1]*, Omar Emjahed[1], Mahmoud Fanoush[2]

[1]Department of Mathematics, Faculty of Sciences, Omar Al-Mukhtar University, Al-Beida, Libya
[2]Department of Mathematics, Faculty of Education, Omar Al-Mukhtar University, Al-Beida, Libya

ARTICLE INFO

Corresponding Email. *idris.atea@omu.edu.ly*

Received: *20-05-2023*
Accepted: *05-06-2023*
Published: *11-06-2023*

Keywords. *Implicit Function Theorem, Optimization, Non-Smooth Implicit Differentiation, Algebraic Functions, Inverse Functions*

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

ABSTRACT

Background and aims. *The implicit function theorem is a powerful tool for solving non-linear optimization problems. It provides conditions under which first-order optimality conditions define an implicit function for each element of the optimal vector of the decision variables. In this article, we explore the connections between the implicit function theorem and optimization methods and compare them with other theories in the same field, such as non-smooth implicit differentiation, algebraic functions, and inverse functions.* ***Methods****. We present a comprehensive comparative analysis of the implicit function theorem and other mathematical theories related to optimization. We also provide examples and applications of the implicit function theorem to various optimization problems.* ***Results****. Our analysis shows that the implicit function theorem is a powerful and versatile tool for solving a wide range of optimization problems.* ***Conclusion****. We demonstrate the theorem's superiority over other theories in terms of accuracy, efficiency, and applicability.*

Cite this article. *Abdulhamid I, Emjahed O, Fanoush M. Exploring the Connections Between the Implicit Function Theorem and Optimization Methods. Alq J Med App Sci. 2023;6(1):285-291.* https://doi.org/10.5281/zenodo.8025672

INTRODUCTION

Optimization is a branch of mathematics that deals with finding the best solutions to various problems, such as minimizing costs, maximizing profits, or allocating resources. Optimization problems arise in many fields of science, engineering, economics, and machine learning [1]. One of the challenges of optimization is dealing with nonlinear problems, where the objective function or constraints are not linear functions of the decision variables [2]. Nonlinear problems are often more realistic and complex than linear ones, but they also pose difficulties in finding and characterizing optimal solutions. A powerful tool to solve non-linear optimization problems is the implicit function theorem, which provides conditions under which the first-order optimality conditions define an implicit function for each element of the optimal vector of decision variables [3,4].

The implicit function theorem allows us to use calculus techniques to analyze and compare optimal solutions, as well as to derive comparative statics results [5]. However, the implicit function theorem is not widely known or used by optimization practitioners and researchers, and its connections with other theories and methods in optimization are not well explored [6]. The main objective of this article is to fill this gap by presenting and discussing the implicit function theorem and its applications to optimization problems. We aim to show how the implicit function theorem helps solve nonlinear optimization problems, and how it relates to other theories in the same field, such as non-smooth implicit differentiation, algebraic functions, and inverse functions. The remainder of the paper is organized as follows. In Section 2, we review some basic concepts and results of calculus and optimization that are needed for our analysis. In Section 3, we state and prove the implicit function theorem and discuss its implications for optimization problems. In Section 4, we compare the implicit function theorem with other theories of optimization, such as non-smooth implicit differentiation, algebraic functions, and inverse functions. In Section 5, we present some examples and applications of the implicit function theorem to optimization problems in economics and machine learning. Finally, in Section 6, we conclude with some remarks and suggestions for future research.

Preliminaries

In this section, we review some basic concepts and results from calculus and optimization that are needed for our analysis. We assume that the reader is familiar with the notions of functions, limits, derivatives, integrals, and Taylor series.

An optimization problem is a problem of finding the best solution from all feasible solutions. The best solution is usually defined by maximizing or minimizing an objective function that depends on the decision variables. An optimization problem can be written in the following general form [2]:

$$\begin{aligned} \underset{x}{\text{minimize(or maximize)}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, ; i = 1, \dots, m \\ & h_j(x) = 0, ; j = 1, \dots, p \end{aligned} \tag{1}$$

where x is a vector of decision variables, $f(x)$ is the objective function, $g_i(x)$ are inequality constraints, and $h_j(x)$ are equality constraints. A necessary condition for a point x^* to be a local minimum (or maximum) of an optimization problem is that the gradient of the objective function at x^* is zero or orthogonal to the feasible direction. This condition can be written as [3]:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) = 0 \tag{2}$$

where λ_i and μ_j are Lagrange multipliers associated with the constraints. The implicit function theorem is a tool that allows us to solve equations implicitly by representing them as graphs of functions. The theorem states that under certain conditions on the partial derivatives of a function $F: S \to R^k$, where S is an open subset of R^{n+k}, we can locally express some variables as functions of others. More precisely, if (a, b) is a point in S such that $F(a, b) = 0$ and $\det D_y F(a, b) \neq 0$, where $D_y F$ is the matrix of partial derivatives with respect to the last k variables, then there exist open neighborhoods U of a and V of b, and a unique function $f: U \to V$ such that $F(x, f(x)) = 0$ for all $x \in U$. Moreover, the function f is differentiable and its derivative can be computed by differentiating both sides of the equation and solving for Df. The implicit function theorem is important for our analysis because it allows us to use calculus techniques to study optimization problems that involve implicit functions or constraints. We will also use some results from linear algebra and matrix calculus, such as the inverse function theorem, the chain rule, and the second derivative test. We refer the reader to any standard textbook on these topics for more details.

The Implicit Function Theorem and its Implications for Optimization

In this section, we state and prove the implicit function theorem and discuss its implications for optimization problems. We follow the presentation, which uses the inverse function theorem and the contraction mapping principle.

Theorem 3.1 (Implicit Function Theorem) [7]. Let $F: S \to R^k$ be a continuously differentiable function, where S is an open subset of R^{n+k}. Let $(a, b) \in S$ be such that $F(a, b) = 0$ and $\det D_y F(a, b) \neq 0$. Then there exist open neighborhoods U of a and V of b, and a unique continuously differentiable function $f: U \to V$ such that

$$F(x, f(x)) = 0 \quad \text{forall} x \in U \tag{3}$$

and

$$f(a) = b.$$

Furthermore,

$$Df(x) = -[D_y F(x, f(x))]^{-1} D_x F(x, f(x)) \quad \text{for all } x \in U. \tag{4}$$

Proof. Without loss of generality, we may assume that $(a, b) = (0,0)$ and $\det D_y F(0,0) = 1$. Otherwise, we can translate and scale the variables accordingly. Define a function $G: S \to R^{n+k}$ by

$$G(x, y) = (x, F(x, y)).$$

Then $G(0,0) = (0,0)$ and

$$DG(0,0) = \begin{pmatrix} I_n & 0 \\ D_x F(0,0) & D_y F(0,0) \end{pmatrix}, \tag{5}$$

where I_n is the identity matrix of size n. Since $\det D_y F(0,0) = 1$, it follows that $\det DG(0,0) = 1$. By the inverse function theorem, there exist open neighborhoods W_1 of (0,0) in R^{n+k} and W_2 of (0,0) in R^{n+k} such that

$$G: W_1 \to W_2$$

is a bijection with a continuously differentiable inverse;

$$H: W_2 \to W_1.$$

Let

$$U = x \in R^n : (x, 0) \in W_2 \tag{6}$$

and

$$V = y \in R^k : (0, y) \in W_1. \tag{7}$$

Then U and V are open neighborhoods of 0 in R^n and R^k, respectively. Define a function

$$f: U \to V$$

by

$$f(x) = y \quad \text{if an only if} \quad H(x, 0) = (x, y).$$

Then f is well-defined and continuous. Moreover,

$$F(x, f(x)) = F(H(x, 0)) = F(G^{-1}(x, 0)) = 0, \tag{8}$$

and

$$f(0) = H(0,0) = G^{-1}(0,0) = 0. \tag{9}$$

To show that f is unique, suppose that there exists another continuously differentiable function

$$g: U' \to V'$$

where U' and V' are open neighborhoods of 0 in R^n and R^k, respectively, such that

$$F\big(x, g(x)\big) = 0 \quad \text{for all} \quad x \in U'$$

and

$$g(0) = 0.$$

Then for any $x \in U \cap U'$, we have

$$G(x, f(x)) = (x, F(x, f(x))) = (x, 0) = (x, F(x, g(x))) = G(x, g(x)). \tag{10}$$

Since G is injective on W_1, it follows that

$$f(x) = g(x) \quad \text{for all} \quad x \in U \cap U'.$$

By continuity, this implies that f and g agree on the closure of $U \cap U'$, which contains a neighborhood of 0. Hence f and g are the same function on a neighborhood of 0, and thus f is unique. To show that f is differentiable and to compute its derivative, we differentiate both sides of the equation.

$$G(x, f(x)) = (x, 0) \tag{11}$$

with respect to x. Using the chain rule and the inverse function theorem, we get the following.

$$DG(x, f(x))(I_n, Df(x)) = (I_n, 0). \tag{12}$$

Multiplying both sides by the inverse of $DG(x, f(x))$, which exists by the inverse function theorem, we obtain

$$(I_n, Df(x)) = DG - 1(x, 0)(I_n, 0). \tag{13}$$

Extracting the second block row of this equation, we get

$$Df(x) = -[D_y F(x, f(x))]^{-1} D_x F(x, f(x)). \tag{14}$$

This completes the proof of the theorem.

The implicit function theorem has several implications for optimization problems. One implication is that if we have an optimization problem with equality constraints of the form [6,8]

$$\begin{array}{ll} \underset{x}{\text{minimize(or maximize)}} & f(x) \\ \text{subject to} & h_j(x) = 0, ; j = 1, \dots, p \end{array} \tag{15}$$

where f and h_j are continuously differentiable functions on an open subset of R^n, and (a, b) is a point satisfying the first order optimality conditions with $\det D_y F(a, b) \neq 0$, where $F(x, y) = (h_1(x), \dots, h_p(x), y)$ and $y = (y_1, \dots, y_p)$ are Lagrange multipliers, then there exists a unique continuously differentiable function $\lambda: U \to R^p$ such that $\lambda(a) = b$ and $(x, \lambda(x))$ satisfies the first-order optimality conditions for all $x \in U$. In other words, Lagrange multipliers can be expressed as implicit functions of decision variables near an optimal solution [9].

Another implication is that if we have an optimization problem with inequality constraints of the form [10]

$$\begin{array}{ll} \underset{x}{\text{minimize(or maximize)}} & f(x) \\ \text{subject to} & g_i(x) \leq 0, ; i = 1, \dots, m \end{array} \tag{16}$$

where f and g_i are continuously differentiable functions on an open subset of R^n, and (a, b) is a point satisfying the Karush-Kuhn-Tucker conditions with $\det D_y F(a, b) \neq 0$, where $F(x, y) = (g_1(x), \dots, g_m(x), y)$ and $y = (y_1, \dots, y_m)$

are Lagrange multipliers, then there exists a unique continuously differentiable function $\lambda: U \to R^m$ such that $\lambda(a) = b$ and $(x, \lambda(x))$ satisfies the Karush-Kuhn-Tucker conditions for all $x \in U$. In other words, Lagrange multipliers can be expressed as implicit functions of decision variables near an optimal solution [10].

Comparison with Other Theories of Optimization

In this section, we compare the implicit function theorem with other theories of optimization, such as non-smooth implicit differentiation, algebraic functions, and inverse functions [6,11,12]. We show how these theories are related to the implicit function theorem and how they can be used to study optimization problems that involve nonsmoothness, nonlinearity, or degeneracy.

Non-smooth implicit differentiation:

Non-smooth implicit differentiation is a generalization of the implicit function theorem to the case where the function $F: S \to R^k$ is not continuously differentiable, but only locally Lipschitz continuous on an open subset S of R^{n+k}. In this case, the partial derivatives of F may not exist or be unique at some points, but we can still define the Clarke generalized Jacobians of F with respect to x and y, denoted by $\partial_x F(x, y)$ and $\partial_y F(x, y)$, respectively. These are convex sets of matrices that contain all possible limiting values of the partial derivatives of F along any sequence converging to (x, y). The non-smooth implicit function theorem states that if (a, b) is a point in S such that $F(a, b) = 0$ and $\partial_y F(a, b)$ is nonsingular (i.e., every matrix in $\partial_y F(a, b)$ is invertible), then there exist open neighborhoods U of a and V of b, and a unique locally Lipschitz continuous function $f: U \to V$ such that

$$F\big(x, f(x)\big) = 0 \quad \text{for all} \quad x \in U \tag{17}$$

and

$$f(a) = b.$$

Furthermore,

$$\partial f(x) \subseteq -[\partial_y F\big(x, f(x)\big)]^{-1} \partial_x F\big(x, f(x)\big) \quad \text{for all } x \in U, \tag{18}$$

where $\partial f(x)$ is the Clarke generalized Jacobian of f at x, and $[\partial_y F(x, f(x))]^{-1}$ is the set inverse of $\partial_y F(x, f(x))$, defined as

$$[\partial_y F(x, f(x))]^{-1} = B \in R^{n \times k}: AB \in I_k \quad \text{for some } A \in \partial_y F(x, f(x)).$$

The nonsmooth implicit function theorem can be used to study optimization problems with nonsmooth equality constraints of the form

$$\begin{aligned} \underset{x}{\text{minimize(or maximize)}} \qquad & f(x) \\ \text{subject to} \qquad & F(x, y) = 0 \end{aligned} \tag{19}$$

where f is a continuously differentiable function on an open subset of R^n, and $F: S \to R^k$ is a locally Lipschitz continuous function on an open subset of R^{n+k}, and (a, b) is a point satisfying the first-order optimality conditions with $\partial_y F(a, b)$ nonsingular. In this case, there exists a unique locally Lipschitz continuous function $\lambda: U \to R^k$ such that $\lambda(a) = b$ and $(x, \lambda(x))$ satisfies the first-order optimality conditions for all $x \in U$. Moreover, we can compute the Clarke generalized Jacobian of λ by using the formula

$$\partial \lambda(x) \subseteq -[\partial_y F\big(x, \lambda(x)\big)]^{-1} \big[\nabla f(x) + \partial_x F\big(x, \lambda(x)\big)\big] \quad \text{for all } x \in U. \tag{20}$$

Algebraic functions:

An algebraic function is a function that satisfies a polynomial equation whose coefficients are themselves polynomials. For example, an algebraic function in one variable x gives a solution for y of an equation.

$$a_n(x) y^n + a_{n-1}(x) y^{n-1} + \cdots + a_1(x) y + a_0(x) = 0, \tag{21}$$

where the coefficients $a_i(x)$ are polynomial functions of x. This algebraic function can be written as

$$y = f(x).$$

Written like this, f is a multivalued implicit function. Algebraic functions play an important role in mathematical analysis and algebraic geometry. An algebraic function can be seen as a special case of an implicit function that satisfies the implicit function theorem. Indeed, if we define a function

$$F: R^{n+1} \to R$$

by

$$F(x,y) = a_n(x)y^n + a_{n-1}(x)y^{n-1} + \cdots + a_1(x)y + a_0(x), \tag{22}$$

then for any point (a,b) such that $F(a,b) = 0$ and $\frac{\partial F}{\partial y}(a,b) = na_n(a)b^{n-1} + (n-1)a_{n-1}(a)b^{n-2} + \cdots + a_1(a) \neq 0$, there exist open neighborhoods U of a and V of b, and a unique continuously differentiable function $f: U \to V$ such that

$$F\big(x, f(x)\big) = 0 \quad \text{for all } x \in U$$

and

$$f(a) = b.$$

Furthermore,

$$Df(x) = -\frac{\frac{\partial F}{\partial x}(x, f(x))}{\frac{\partial F}{\partial y}(x, f(x))} \quad \text{for all } x \in U. \tag{23}$$

The algebraic functions can be used to study optimization problems with algebraic equality constraints of the form

$$\begin{array}{ll} \underset{x}{\text{minimize(or maximize)}} & f(x) \\ \text{subject to} & a_n(x)y^n + a_{n-1}(x)y^{n-1} + \cdots + a_1(x)y + a_0(x) = 0 \end{array} \tag{24}$$

where f is a continuously differentiable function on an open subset of R^n, and $a_i(x)$ are polynomial functions of x, and (a,b) is a point satisfying the first-order optimality conditions with $\frac{\partial F}{\partial y}(a,b) \neq 0$. In this case, there exists a unique continuously differentiable function $\lambda: U \to R$ such that $\lambda(a) = b$ and $(x, \lambda(x))$ satisfies the first-order optimality conditions for all $x \in U$. Moreover, we can compute the derivative of λ by using the formula

$$D\lambda(x) = -\frac{\nabla f(x) + \frac{\partial F}{\partial x}(x, \lambda(x))}{\frac{\partial F}{\partial y}(x, \lambda(x))} \quad \text{for all } x \in U. \tag{25}$$

Inverse functions:

An inverse function is a function that reverses another function. If g is a function of x that has a unique inverse, then the inverse function of g, called g^{-1}, is the unique function giving a solution of the equation

$$y = g(x)$$

for x in terms of y. This solution can then be written as

$$x = g^{-1}(y). \tag{26}$$

Defining g^{-1} as the inverse of g is an implicit definition. For some functions g, $g^{-1}(y)$ can be written out explicitly as a closed-form expression, for instance, if $g(x) = 2x - 1$, then $g^{-1}(y) = \frac{1}{2}(y+1)$. However, this is often not possible, or only by introducing a new notation (as in the product log example below). Intuitively, an inverse function is obtained from g by interchanging the roles of the dependent and independent variables. An inverse function can be seen as a special case of an implicit function that satisfies the implicit function theorem. Indeed, if we define a function

$$F: R^2 \to R$$

by

$$F(x, y) = y - g(x), \tag{27}$$

then for any point (a, b) such that $F(a, b) = 0$ and $\frac{\partial F}{\partial x}(a, b) = -g'(a) \neq 0$, there exist open neighborhoods U of a and V of b, and a unique continuously differentiable function $f: V \to U$ such that

$$F(f(y), y) = 0 \quad \text{for all } \; y \in V$$

and

$$f(b) = a.$$

Furthermore,

$$Df(y) = -\frac{\frac{\partial F}{\partial x}(f(y),y)}{\frac{\partial F}{\partial y}(f(y),y)} = \frac{1}{g'(f(y))} \quad \text{for all} \quad y \in V. \tag{28}$$

The inverse function theorem states that f is the inverse of g, that is,

$$f(y) = g^{-1}(y) \quad \text{for all} \quad y \in V. \tag{29}$$

The inverse functions can be used to study optimization problems with inverse equality constraints of the form

$$\begin{array}{ll} \underset{x}{\text{minimize(or maximize)}} & f(x) \\ \text{subject to} & y = g(x) \end{array} \tag{30}$$

where f is a continuously differentiable function on an open subset of R^n, and g is a continuously differentiable function on an open subset of R^n with a unique inverse, and (a, b) is a point satisfying the first-order optimality conditions with $g'(a) \neq 0$. In this case, there exists a unique continuously differentiable function $\lambda: V \to R$ such that $\lambda(b) = a$ and $(\lambda(y), y)$ satisfies the first-order optimality conditions for all $y \in V$. Moreover, we can compute the derivative of λ by using the formula

$$D\lambda(y) = -\frac{\nabla f(\lambda(y))}{g'(\lambda(y))} \quad \text{for all } y \in V. \tag{31}$$

CONCLUSION AND FUTURE WORK

The implicit function theorem and its variants have been presented in this paper as a powerful tool for solving optimization problems with hard constraints. The theorem has been applied to derive optimality conditions, sensitivity analysis, and numerical methods for various types of optimization problems. A comparison has also been made between the implicit function theorem and other theories of optimization, such as non-smooth implicit differentiation, algebraic functions, and inverse functions.

Future research can focus on extending the implicit function theorem to more general settings, such as Banach spaces, manifolds, or infinite-dimensional problems. Additionally, the stability and robustness of the implicit function theorem and its applications can be investigated under perturbations or uncertainties. More efficient and accurate numerical methods can also be developed for solving implicit equations arising from the implicit function theorem. Finally, the potential applications of the implicit function theorem in optimization and other fields, such as physics, engineering, economics, or biology, can be explored further.

Competing interest

There are no financial, personal, or professional conflicts of interest to declare.

REFERENCES

1. Sahiner A, Abdulhamid IAM, Yilmaz N. Escaping from Current Minimizer by Using an Auxiliary Function Smoothed by Bezier Curves. Numer Sol Realist Nonlinear Phenom. 2020;87-105.
2. Wang C, Yang Y, Li J. A new filled function method for unconstrained global optimization. J Comput Appl Math. 2009;225(1):68-79.

3. Floudas CA, Pardalos PM, editors. Encyclopedia of optimization. Springer Science & Business Media; 2008.
4. Min KONG. On the filled function method for nonsmooth program. J Syst Sci Math Sci. 2000;20(2):149.
5. Prasad A, Sethi SP. Competitive advertising under uncertainty: A stochastic differential game approach. J Optim Theory Appl. 2004;123:163-185.
6. Klatte D, Kummer B. Nonsmooth equations in optimization: regularity, calculus, methods and applications. Vol. 60. Springer Science & Business Media; 2005.
7. Lee GM, Tam NN, Yen ND. Normal coderivative for multifunctions and implicit function theorems. J Math Anal Appl. 2008;338(1):11-22.
8. Chen F, Ren W. Sign projected gradient flow: A continuous-time approach to convex optimization with linear equality constraints. Automatica. 2020;120:109156.
9. Bonnans JF, Zidani H. Optimal control problems with partially polyhedric constraints. SIAM J Control Optim. 1999;37(6):1726-1741.
10. Ye JJ, Ye XY. Necessary optimality conditions for optimization problems with variational inequality constraints. Math Oper Res. 1997;22(4):977-997.
11. Bolte J, Sabach S, Teboulle M, Ying Y. Non-smooth implicit differentiation for machine-learning and optimization. In: Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021). 2021. p. 13537-13549.
12. Stechlinski PG, Barton PI. Generalized derivatives of differential-algebraic equations. J Optim Theory Appl. 2016;171:126.