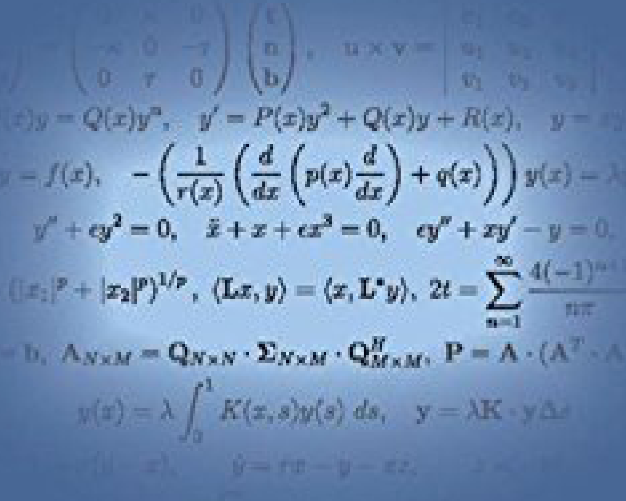


Mathematical Methods in Engineering

Joseph M. Powers • Mihir Sen

MATHEMATICAL METHODS IN ENGINEERING

This text focuses on a variety of topics in mathematics in common usage in graduate engineering programs including vector calculus, linear and nonlinear ordinary differential equations, approximation methods, vector spaces, linear algebra, integral equations, and dynamical systems. The book is designed for engineering graduate students who wonder how much of their basic mathematics will be of use in practice. Following development of the underlying analysis, the book takes students step-by-step through a large number of examples that have been worked in detail. Students can choose to go through each step or to skip ahead if they desire. After seeing all the intermediate steps, they will be in a better position to know what is expected of them when solving homework assignments and examination problems, and when they are on the job. Each chapter concludes with numerous exercises for the student that reinforce the chapter content and help connect the subject matter to a variety of engineering problems. Students today have grown up with computer-based tools including numerical calculations and computer graphics; the worked-out examples as well as the end-of-chapter exercises often use computers for numerical and symbolic computations and for graphical display of the results.

Joseph M. Powers joined the University of Notre Dame in 1989. His research has focused on the dynamics of high-speed reactive fluids and on computational science, especially as it applies to verification and validation of complex multiscale systems. He has held positions at the NASA Lewis Research Center, the Los Alamos National Laboratory, the Air Force Research Laboratory, the Argonne National Laboratory, and the Chinese Academy of Sciences. He is a member of AIAA, APS, ASME, the Combustion Institute, and SIAM. He is the recipient of numerous teaching awards.

Mihir Sen has been active in teaching and in research in thermal-fluids engineering – especially in regard to problems relating to modeling, dynamics, and stability – since obtaining his PhD from MIT. He has worked on reacting flows, natural and forced convection, flow in porous media, falling films, boiling, MEMS, heat exchangers, thermal control, and intelligent systems. He joined the University of Notre Dame in 1986 and received the Kaneb Teaching Award from the College of Engineering in 2001 and the Rev. Edmund P. Joyce, C.S.C., Award for Excellence in Undergraduate Teaching in 2009. He is a Fellow of ASME.

Mathematical Methods in Engineering

Joseph M. Powers
University of Notre Dame

Mihir Sen
University of Notre Dame

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107037045

First published 2015

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data
Powers, Joseph, 1961–
Mathematical methods in engineering / Joseph Powers, University of Notre Dame,
Mihir Sen, University of Notre Dame.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-03704-5 (hardback : alk. paper)
1. Engineering mathematics – Study and teaching (Graduate)   I. Sen, Mihir, 1947–
II. Title.
TA332.5.P69   2015
510–dc23   2014020962

ISBN 978-1-107-03704-5 Hardback

To my parents: Mary Rita, my first reading teacher, and Joseph Leo, my first mathematics teacher – Joseph Michael Powers

To my family: Beatriz, Pradeep, Maya, Yasamin, and Shayan – Mihir Sen

Contents

Preface

Our overarching aim in writing this book is to build a bridge to enable engineers to better traverse the domains of the mathematical and physical worlds. Our focus is on neither the nuances of pure mathematics nor the phenomenology of physical devices but instead is on the mathematical tools used today in many engineering environments. We often compromise strict formalism for the sake of efficient exposition of mathematical tools. Whereas some results are fully derived, others are simply asserted, especially when detailed proofs would significantly lengthen the presentation. Thus, the book emphasizes method and technique over rigor and completeness; readers who require more of the latter can and *should* turn to many of the foundational works cited in the extensive bibliography.

Our specific objective is to survey topics in engineering-relevant applied mathematics, including multivariable calculus, vectors and tensors, ordinary differential equations, approximation methods, linear analysis, linear algebra, linear integral equations, and nonlinear dynamical systems. In short, the text fully explores linear systems and considers some effects of nonlinearity, especially those types that can be treated analytically. Many topics have geometric interpretations, identified throughout the book. Particular attention is paid to the notion of approximation via projection of an entity from a high- or even infinite-dimensional space onto a space of lower dimension. Another goal is to give the student the mathematical background to delve into topics such as dynamics, differential geometry, continuum mechanics, and computational methods; although the material presented is relevant to those fields, specific physical applications are mainly confined to some of the exercises. A final goal is to introduce the engineer to *some* of the notation and rigor of mathematics in a way used in many upper-level graduate engineering and applied mathematics courses.

This book is intended for use in a beginning graduate course in applied mathematics taught to engineers. It arose from a set of notes for such a course taught by the authors for more than 20 years in the Department of Aerospace and Mechanical Engineering at the University of Notre Dame. Students in this course come from a variety of backgrounds, mainly within engineering but also from science. Most enter with some undergraduate-level proficiency in differential and integral multivariable calculus, differential equations, vectors analysis, and linear algebra. This book briefly reviews these subjects but more often builds on an assumed elementary understanding of topics such as continuity, limits, series, and the chain rule.

As such, we often casually introduce subject matter whose full development is deferred to later chapters. For example, although one of the key features of the book is a lengthy discussion of eigensystems in Chapter 6, most engineers will already know what an eigenvalue is. Consequently, we employ eigenvalues in nearly every chapter, starting from Chapter 1. The same can be said for topics such as vector operators, determinants, and linear equations. When such topics are introduced earlier than their formal presentation, we often make a forward reference to the appropriate section, and the student is encouraged to read ahead. In summary, most beginning graduate students and advanced undergraduates will be prepared for the subject matter, though they may find occasion to revisit some trusted undergraduate texts. Although our course is only one semester, we have added some topics to those we usually cover; the instructor of a similar course should be able to omit certain topics and add others.

At a time not very far in the past, mathematics in engineering was largely confined to basic algebra and interpolation of trigonometric and logarithmic tables. Not so today! Much of engineering has come to rely on sophisticated predictive mathematical models to design and control a wide variety of devices, for example, buildings and bridges, air and ground transportation, manufacturing and chemical processes, electrical and electronic devices, and biomedical and robotic equipment. These models may be in the form of algebraic, differential, integral, or other equations. Formulation of such models is often a challenge that calls on the basic sciences of physics, chemistry, and biology. Once they are formulated, the engineer is faced with actually solving the model equations, and for that a variety of tools are of value. Our focus is on the general mathematical tools used for engineering problems but not their formulation or specific physical details. While we sporadically discuss a paradigm problem such as a mass-spring-damper, we focus more on the mathematics. The use of mathematical analysis within engineering has changed greatly over the years. Once it was the sole means to the solution of some problems, but currently engineers rely on it within numerical and experimental approaches. The use of the adjective *numerical* has also changed over the years because of the variety of ways in which computers may be used in engineering. It is in fact becoming more common and necessary for extensive mathematical manipulation to be performed to prepare the computer for efficient and accurate solution generation.

Choices have been made with regard to notation; most of the conventions we adopt are reflected in at least a portion of the literature. In a few cases, we choose to diverge slightly from some of the more common norms. When we do so, explanatory footnotes are usually included. For example, the literature has a number of conventions for the so-called dot product or scalar product between two vectors, $\mathbf{u}$ and $\mathbf{v}$. The most common is probably $\mathbf{u} \cdot \mathbf{v}$. We generally choose the more elaborate $\mathbf{u}^T \cdot \mathbf{v}$, where the T indicates a transpose operation. This emphasizes the fact that vectors are considered to be columns of elements and that to associate a scalar product of two vectors with the ordinary rules of matrix multiplication, one needs to transpose the first vector into a row of elements. Similarly, we generally take the product of a matrix $\mathbf{A}$ and vector $\mathbf{u}$ to be of the form $\mathbf{A} \cdot \mathbf{u}$ rather than the often-used $\mathbf{Au}$. And we use $\mathbf{u}^T \cdot \mathbf{A}$, while many texts simply write $\mathbf{uA}$. Unusually, we often apply the transpose notation to the so-called divergence operator, writing, for example, the divergence of a vector field $\mathbf{u}$ as $\operatorname{div} \mathbf{u} = \nabla^T \cdot \mathbf{u}$ rather than as the more common $\nabla \cdot \mathbf{u}$. One could easily infer the nature of the divergence operation without the transpose, but we believe it adds unity to our notational framework. In the text, italicized letters like a will most often be used for scalars, bold lowercase letters like $\mathbf{a}$ for vectors, bold uppercase letters like $\mathbf{A}$ for tensors and operators, and $\mathbb{A}$ for sets and spaces. In general, we use ^{T} for transpose, $^-$ for complex conjugate, * for adjoint, and ^{H} for Hermitian transpose. The student also has to be aware that the same quantities written on a blackboard or paper may appear differently. Whatever the notational choices, the student should be fluent in a variety of usages in the literature; we in fact sometimes deviate from our own conventions to reinforce that our choices are not unique.

Our experience has been that engineering students learn best by exposure to examples, and a hallmark of the text is that much of the material is developed via presentation of a large number of fully worked problems, each of which generally follows a short fundamental development. The solved examples not only illustrate the points made previously but also introduce additional concepts and are thus an integral part of the text. Ultimately, mathematics is learned by doing, and for this reason we have a large number of exercises. Engineers, moreover, have a special purpose for studying mathematics: they need it to solve practical problems; some are included as exercises.

Presentation of many of our specific details has relied on modern software for symbolic computation and plotting. We encourage the reader to utilize these tools as well, as they enable exact solutions and graphics that may otherwise be impossible. The text does not provide details of particular software packages, which often change with each new version; the reader is advised to choose one or more packages, such as Mathematica, Maple, or MATLAB, and to become familiar with its usage. Exercises are included that require the use of such software. The phrase "symbolic computer mathematics" is used to mean tools such as Mathematica or Maple and "discrete computational methods" to connote tools such as MATLAB, Python, Fortran, C, or C++. The problems emphasize plotting to give a geometric overview of the results. The use of visuals or graphics to get a quick appreciation for the quality of an approximation or the behavior of a result cannot be overemphasized.

There are a number of texts on graduate-level mathematics for engineers. Mathematics applied to engineering is a vast discipline; consequently, each book has a unique emphasis. Here we have attempted to include what is actually used by researchers in our field. To be clear, though, because it is for a one-semester introductory course, many topics are left for advanced courses. Among the topics omitted or lightly covered are integral transforms, complex variables, partial differential equations, group theory, probability, statistics, numerical methods, and graph and network theory.

In closing, we express our hope that the readers of this book will find mathematics to be as beautiful and useful a subject as we have over the years. Our appreciation was nurtured by a large number of special people: family members, teachers at all levels, colleagues at home and abroad, authors from many ages, and our own students over the decades. We have learned from all of them and hope that our propagation of their knowledge engenders new discoveries from readers of this book for future generations.

Joseph M. Powers
Mihir Sen
Notre Dame, Indiana, USA

1 Multivariable Calculus

We commence with a presentation of fundamental notions from the *calculus of many variables*, a subject that will be seen to provide a useful language to describe multidimensional geometry, a topic foundational to nearly all of engineering. We focus on presenting its grammar in sufficient richness to enable precise rendering of both simple and more complex geometries. This is of self-evident importance in engineering applications. We begin with a summary of implicitly defined functions and, in what will be a foundation for the book, local linearization. This quickly leads us to consider matrices that arise from linearization. The basic strategies of optimization, both unconstrained and constrained, are considered via identification of maxima and minima as well as the calculus of variations and Lagrange multipliers. The chapter ends with an extensive exposition relating multivariable calculus to the geometry of nonlinear coordinate transformations; in so doing, we develop geometry-based analytical tools that are used throughout the book. Such tools have relevance in fields ranging from computational mechanics widely used in engineering software to the theory of dynamical systems. Understanding the general nonlinear case also prepares one to better frame the more common geometrical methods associated with orthonormal coordinate transformations studied in Chapter 2.

1.1 Implicit Functions

1.1.1 One Independent Variable

We begin with a consideration of implicit functions of the form

$$f(x, y) = 0. \tag{1.1}$$

Here f is allowed to be a fully nonlinear function of x and y. We take the point (x_0, y_0) to satisfy $f(x_0, y_0) = 0$. Near (x_0, y_0), we can locally linearize via the familiar *total differential*,

$$df = \left.\frac{\partial f}{\partial x}\right|_{x_0,y_0} dx + \left.\frac{\partial f}{\partial y}\right|_{x_0,y_0} dy = 0. \tag{1.2}$$

At the point (x_0, y_0), both $\partial f/\partial x$ and $\partial f/\partial y$ are constants, and we consider the local "variables" to be dx and dy, the infinitesimally small deviations of x and y from x_0

and y_0. As long as $\partial f/\partial y \neq 0$, we can scale to say

$$\left.\frac{dy}{dx}\right|_{x_0,y_0} = -\frac{\left.\dfrac{\partial f}{\partial x}\right|_{x_0,y_0}}{\left.\dfrac{\partial f}{\partial y}\right|_{x_0,y_0}}. \tag{1.3}$$

And with the local existence of dy/dx, we can consider y to be a unique local linear function of x and quantify its variation as

$$y - y_0 \approx \left.\frac{dy}{dx}\right|_{x_0,y_0}(x - x_0). \tag{1.4}$$

The process of examining the derivative at a point is a fundamental part of the use of linearization to gain local knowledge of a nonlinear entity. From here on we generally do not specify the point at which the evaluation is occurring, as it should be easy to understand from the context of the problem.

These notions can be formalized in the following:

Implicit Function Theorem: For a given $f(x, y)$ with $f = 0$ and $\partial f/\partial y \neq 0$ at the point (x_0, y_0), there corresponds a unique function $y(x)$ in the neighborhood[1] of (x_0, y_0).

When such a condition is satisfied, we can consider the dependent variable y to be locally a unique function of the independent variable x. The theorem says nothing about the case $\partial f/\partial y = 0$. Though such a case explicitly states that at such a point f is not a function of y, it does not speak to $y(x)$.

EXAMPLE 1.1

Consider the implicit function

$$f(x, y) = x^2 - y = 0. \tag{1.5}$$

Determine where y can be considered to be a unique local function of x.

We first recognize that when $f = 0$, we have $y = x^2$, which is a simple parabola. And it is easy to see that for all $x \in (-\infty, \infty)$, y is defined[2] and will be such that $y \in [0, \infty)$. Let us apply the formalism of the implicit function theorem to see how it is exercised on this well-understood curve. Applying Eq. (1.2) to our f, we find

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = 0, \tag{1.6}$$

$$= 2x\,dx - \,dy = 0. \tag{1.7}$$

Thus,

$$\frac{dy}{dx} = 2x. \tag{1.8}$$

This is defined for all $x \in (-\infty, \infty)$. Thus, a unique linearization $y(x)$ exists for all $x \in (-\infty, \infty)$. In the neighborhood of a point (x_0, y_0), we have the local linearization

$$y - y_0 = 2x_0(x - x_0). \tag{1.9}$$

[1] *Neighborhood* is a word whose mathematical meaning is close to its common meaning; nevertheless, it does require a precise definition, which is deferred until Section 6.1.

[2] Here $\in$ stands for "is an element of" and indicates that x could be any number, which we assume to be real. Also, we use the standard notation $(\cdot)$ to describe an open interval on the real axis, that is, one that does not include the endpoints. A closed interval is denoted by $[\cdot]$ and mixed intervals by either $(\cdot]$ or $[\cdot)$.

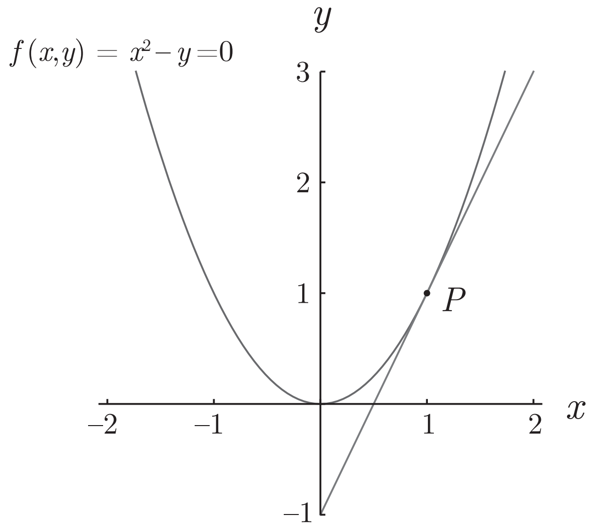


Figure 1.1. The parabola defined implicitly by $f(x, y) = x^2 - y = 0$ along with its local linearization near $P : (1, 1)$.

Consider the point $P : (x_0, y_0) = (1, 1)$. Obviously at P, $f = 1^2 - 1 = 0$, so P is on the curve. The local linearization at P is

$$y - 1 = 2(x - 1), \tag{1.10}$$

which shows uniquely how y varies with x in the neighborhood of P. The parabola defined by $f(x, y) = x^2 - y = 0$ and the linearization near P are plotted in Figure 1.1.

EXAMPLE 1.2

Consider the implicit function

$$f(x, y) = x^2 + y^2 - 1 = 0. \tag{1.11}$$

Determine where y can be considered to be a unique local function of x.

We first note that f is obviously nonlinear; our analysis will hinge on its local linearization. The function can be rewritten as $x^2 + y^2 = 1$, which is recognized as a circle centered at the origin with radius of unity. By inspection, we will need $x \in [-1, 1]$ to have a real value of y. Let us see how the formalism of the implicit function theorem applies. We have $\partial f/\partial x = 2x$ and $\partial f/\partial y = 2y$. We are thus concerned that wherever $y = 0$, there is no unique function $y(x)$ that can be defined in the neighborhood of such a point. Applying Eq. (1.2) to our f, we find

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = 0, \tag{1.12}$$

$$= 2x\,dx + 2y\,dy = 0. \tag{1.13}$$

Thus,

$$\frac{dy}{dx} = -\frac{x}{y}. \tag{1.14}$$

Obviously, whenever $y = 0$, the derivative dy/dx is undefined, and there is no local linearization. In the neighborhood of a point (x_0, y_0), with $y_0 \neq 0$, we have the local linearization

$$y - y_0 = -\frac{x_0}{y_0}(x - x_0). \tag{1.15}$$

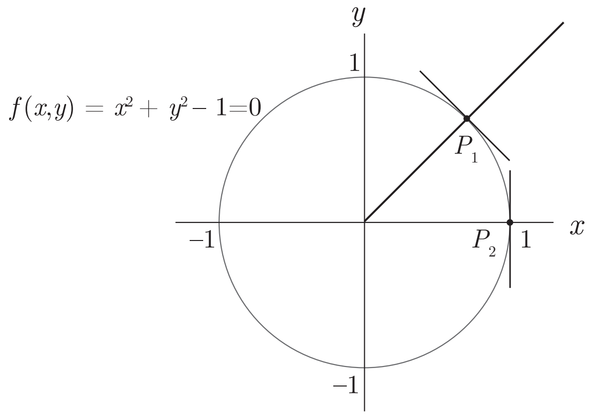


Figure 1.2. The circle defined implicitly by $f(x, y) = x^2 + y^2 - 1 = 0$ along with its local linearization near $P_1 : (\sqrt{2}/2, \sqrt{2}/2)$ and $P_2 : (1, 0)$.

Thus, while y exists for $x \in [-1, 1]$, we only have a local linearization for $x \in (-1, 1)$. When $x = \pm 1$, $y = 0$, and the local linearization does not exist.

Consider two points on the circle $P_1 : (x_0, y_0) = (\sqrt{2}/2, \sqrt{2}/2)$ and $P_2 : (x_0, y_0) = (1, 0)$. Obviously $f = 0$ at both P_1 and P_2, so both are on the circle. At P_1, we have the local linearization

$$y - \frac{\sqrt{2}}{2} = -\left(x - \frac{\sqrt{2}}{2}\right), \tag{1.16}$$

showing uniquely how y varies with x near P_1, the curve being the tangent line to f. At P_2, it is obvious that the tangent line is expressed by $x = 1$, which does not tell us how y varies with x locally. Indeed, we do know that $y = 0$ at P_2, but we lack a unique formula for its variation with x in nearby regions.

The circle defined by $f(x, y) = x^2 + y^2 - 1 = 0$ and the linearizations near P_1 and P_2 are plotted in Figure 1.2. Globally, of course, y does vary with x for this f; in fact, one can see by inspection that

$$y(x) = \pm\sqrt{1 - x^2}. \tag{1.17}$$

This function takes on real values for $x \in [-1, 1]$, but within that domain, it is nowhere unique. One might think a Taylor[3] series (see Section 5.1.1) approximation of Eq. (1.17) might be of use near P_2, but that too is nonunique, yielding the nonlinear approximations

$$y(x) \approx \sqrt{2(1 - x)}, \tag{1.18}$$

$$y(x) \approx -\sqrt{2(1 - x)}, \tag{1.19}$$

for $x \approx 1$. So the impact of the implicit function theorem on this f at $P_2 : (1, 0)$ is that there is no *unique* approximating linear function $y(x)$ at P_2. Whenever a linear approximation is available, we can expect uniqueness.

Had our global function been $f(x, y) = x - 1 = 0$, we would have had $\partial f/\partial y = 0$ everywhere and would nowhere have been able to identify $y(x)$; this simply reflects that for $f = x - 1 = 0$, whenever $x = 1$, any value of y will allow satisfaction of $f = 0$.

[3] Brook Taylor, 1685–1731, English mathematician, musician, and painter.

1.1.2 Many Independent Variables

If, instead, we have N independent variables, we can think of a relation such as

$$f(x_1, x_2, \ldots, x_N, y) = 0 \tag{1.20}$$

in some region as an implicit function of y with respect to the other variables. We cannot have $\partial f/\partial y = 0$, because then f would not depend on y in this region. In principle, we can write

$$y = y(x_1, x_2, \ldots, x_N) \tag{1.21}$$

if $\partial f/\partial y \neq 0$.

The derivative $\partial y/\partial x_n$ can be determined from $f = 0$ without explicitly solving for y. First, from the definition of the total differential, we have

$$df = \frac{\partial f}{\partial x_1}\,dx_1 + \frac{\partial f}{\partial x_2}\,dx_2 + \cdots + \frac{\partial f}{\partial x_n}\,dx_n + \cdots + \frac{\partial f}{\partial x_N}\,dx_N + \frac{\partial f}{\partial y}\,dy = 0. \tag{1.22}$$

Differentiating with respect to x_n while holding all the other $x_m, m \neq n$, constant, we get

$$\frac{\partial f}{\partial x_n} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x_n} = 0, \tag{1.23}$$

so that

$$\frac{\partial y}{\partial x_n} = -\frac{\dfrac{\partial f}{\partial x_n}}{\dfrac{\partial f}{\partial y}}, \tag{1.24}$$

which can be found if $\partial f/\partial y \neq 0$. That is to say, y can be considered a function of x_n if $\partial f/\partial y \neq 0$. Again, local linearization of a nonlinear function provides a key to understanding the behavior of a function in a small neighborhood.

EXAMPLE 1.3

Consider the implicit function

$$f(x_1, x_2, y) = x_1^2 + x_2^2 + y^2 - 1 = 0. \tag{1.25}$$

Determine where y can be considered to be a unique local function of x_1 and x_2.

This problem is essentially a three-dimensional extension of the previous example. Here the nonlinear $f = 0$ describes a sphere centered at the origin with radius unity: $x_1^2 + x_2^2 + y^2 = 1$. We have $\partial f/\partial x_1 = 2x_1$, $\partial f/\partial x_2 = 2x_2$, and $\partial f/\partial y = 2y$. We thus do not expect a unique $y(x_1, x_2)$ to describe the points on the sphere in the neighborhood of $y = 0$.

Applying Eq. (1.22) to our f, we find

$$df = 2x_1\,dx_1 + 2x_2\,dx_2 + 2y\,dy = 0. \tag{1.26}$$

Holding x_2 constant, we can say

$$\frac{\partial y}{\partial x_1} = -\frac{x_1}{y}. \tag{1.27}$$

Holding x_1 constant, we can say

$$\frac{\partial y}{\partial x_2} = -\frac{x_2}{y}. \tag{1.28}$$

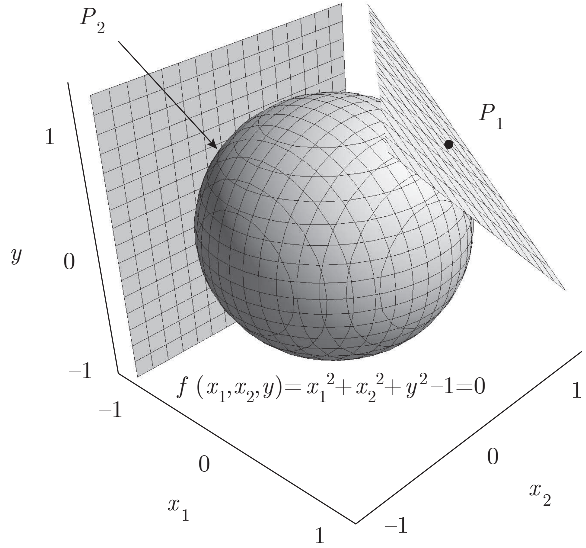


Figure 1.3. The sphere defined implicitly by $f(x_1, x_2, y) = x_1^2 + x_2^2 + y^2 - 1 = 0$ along with its local tangent planes at P_1 : $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ and $P_2 : (-1, 0, 0)$.

Near an arbitrary point (x_{1a}, x_{2a}, y_a) on the surface $f = 0$, we get the equation for the *tangent plane*:

$$y - y_a = -\frac{x_{1a}}{y_a}(x_1 - x_{1a}) - \frac{x_{2a}}{y_a}(x_2 - x_{2a}). \tag{1.29}$$

At the point $P_1 : (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, we can form the unique local linearization

$$y - \frac{1}{\sqrt{3}} = -\left(x_1 - \frac{1}{\sqrt{3}}\right) - \left(x_2 - \frac{1}{\sqrt{3}}\right), \tag{1.30}$$

which describes a tangent plane to $f = 0$ at P_1. At the point $P_2 : (-1, 0, 0)$, the equation

$$x_1 = -1 \tag{1.31}$$

describes the tangent plane. But a unique approximation $y(x_1, x_2)$ for the neighborhood of P_2 does not exist on $f = 0$ at P_2 because $y = 0$ there. Of course, a nonunique approximation does exist, that being $y = \pm\sqrt{1 - x_1^2 - x_2^2}$. The sphere defined by $f(x_1, x_2, y) = x_1^2 + x_2^2 + y^2 - 1 = 0$ along with linearizations near P_1 and P_2 is plotted in Figure 1.3.

1.1.3 Many Dependent Variables

Let us now consider the equations

$$f(x, y, u, v) = 0, \tag{1.32}$$

$$g(x, y, u, v) = 0. \tag{1.33}$$

Under certain circumstances, we can unravel Eqs. (1.32, 1.33), either algebraically or numerically, to form $u = u(x, y)$, $v = v(x, y)$. Those circumstances are given by the multivariable extension of the implicit function theorem. We present it here for two functions of two variables; it can be extended to N functions of N variables.

Theorem: Let $f(x, y, u, v) = 0$ and $g(x, y, u, v) = 0$ be satisfied by (x_0, y_0, u_0, v_0), and suppose that continuous derivatives of f and g exist in the neighborhood of

(x_0, y_0, u_0, v_0) with

$$\left| \begin{array}{cc} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{array} \right|_{x_0,y_0,u_0,v_0} \neq 0. \tag{1.34}$$

Then Eq. (1.34) implies the local existence of $u(x,y)$ and $v(x,y)$.

We will not formally prove this but the following analysis explains its origins. Here we might imagine the two dependent variables u and v both to be functions of the two independent variables x and y. The conditions for the existence of such a functional dependency can be found by local linearized analysis, namely, through differentiation of the original equations. For example, differentiating Eq. (1.32) gives

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial u}\,du + \frac{\partial f}{\partial v}\,dv = 0. \tag{1.35}$$

Holding y constant and dividing by dx, we get

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} = 0. \tag{1.36}$$

Operating on Eq. (1.33) in the same manner, we get

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial x} = 0. \tag{1.37}$$

Similarly, holding x constant and dividing by dy, we get

$$\frac{\partial f}{\partial y} + \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} = 0, \tag{1.38}$$

$$\frac{\partial g}{\partial y} + \frac{\partial g}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial y} = 0. \tag{1.39}$$

The linear Eqs. (1.36, 1.37) can be solved for $\partial u/\partial x$ and $\partial v/\partial x$, and the linear Eqs. (1.38, 1.39) can be solved for $\partial u/\partial y$ and $\partial v/\partial y$ using the well-known Cramer's[4] rule; see Eq. (7.95) or Section A.2. To solve for $\partial u/\partial x$ and $\partial v/\partial x$, we first write Eqs. (1.36, 1.37) in matrix form:

$$\begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{pmatrix} = \begin{pmatrix} -\frac{\partial f}{\partial x} \\ -\frac{\partial g}{\partial x} \end{pmatrix}. \tag{1.40}$$

Thus, from Cramer's rule, we have[5]

$$\frac{\partial u}{\partial x} = \frac{\left| \begin{array}{cc} -\frac{\partial f}{\partial x} & \frac{\partial f}{\partial v} \\ -\frac{\partial g}{\partial x} & \frac{\partial g}{\partial v} \end{array} \right|}{\left| \begin{array}{cc} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{array} \right|} \equiv -\frac{\frac{\partial(f,g)}{\partial(x,v)}}{\frac{\partial(f,g)}{\partial(u,v)}}, \qquad \frac{\partial v}{\partial x} = \frac{\left| \begin{array}{cc} \frac{\partial f}{\partial u} & -\frac{\partial f}{\partial x} \\ \frac{\partial g}{\partial u} & -\frac{\partial g}{\partial x} \end{array} \right|}{\left| \begin{array}{cc} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{array} \right|} \equiv -\frac{\frac{\partial(f,g)}{\partial(u,x)}}{\frac{\partial(f,g)}{\partial(u,v)}}. \tag{1.41}$$

[4] Gabriel Cramer, 1704–1752, Swiss-born mathematician.

[5] Here we are defining notation such as $\partial(f,v)/\partial(u,v)$ in terms of the determinant of quantities, which is easily inferred from the context shown. The notation "$\equiv$" stands for "is defined as."

In a similar fashion, we can form expressions for $\partial u/\partial y$ and $\partial v/\partial y$:

$$\frac{\partial u}{\partial y} = \frac{\begin{vmatrix} -\frac{\partial f}{\partial y} & \frac{\partial f}{\partial v} \\ -\frac{\partial g}{\partial y} & \frac{\partial g}{\partial v} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix}} \equiv -\frac{\frac{\partial(f,g)}{\partial(y,v)}}{\frac{\partial(f,g)}{\partial(u,v)}}, \qquad \frac{\partial v}{\partial y} = \frac{\begin{vmatrix} \frac{\partial f}{\partial u} & -\frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial u} & -\frac{\partial g}{\partial y} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix}} \equiv -\frac{\frac{\partial(f,g)}{\partial(u,y)}}{\frac{\partial(f,g)}{\partial(u,v)}}. \quad (1.42)$$

So if these derivatives locally exist, we can form $u(x, y)$ and $v(x, y)$ locally as well. Alternatively, the same procedure has value in simply finding the derivatives themselves.

We take the *Jacobian*[6] *matrix* $\mathbf{J}$ to be defined for this problem as

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{pmatrix}. \quad (1.43)$$

It is distinguished from the *Jacobian determinant*, J, defined for this problem as

$$J = \det \mathbf{J} = \frac{\partial(f, g)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix} = \frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}. \quad (1.44)$$

If $J \neq 0$, the derivatives exist, and we indeed can find $u(x, y)$ and $v(x, y)$, at least in a locally linear approximation. This is the condition for existence of implicit to explicit function conversion. If $J = 0$, the analysis is not straightforward; typically one must examine problems on a case-by-case basis to make definitive statements.

Let us now consider a simple globally linear example to see how partial derivatives may be calculated for implicitly defined functions.

EXAMPLE 1.4

If

$$x + y + u + v = 0, \quad (1.45)$$

$$2x - y + u - 3v = 0, \quad (1.46)$$

find $\partial u/\partial x$.

We have four unknowns in two equations. Here the problem is sufficiently simple that we can solve for $u(x, y)$ and $v(x, y)$ and then determine all partial derivatives, such as the one desired. Direct solution of the linear equations reveals that

$$u(x, y) = -\frac{5}{4}x - \frac{1}{2}y, \quad (1.47)$$

$$v(x, y) = \frac{1}{4}x - \frac{1}{2}y. \quad (1.48)$$

The surface $u(x, y)$ is a plane and is plotted in Figure 1.4. We could generate a similar plot for $v(x, y)$. We can calculate $\partial u/\partial x$ by direct differentiation of Eq. (1.47):

$$\frac{\partial u}{\partial x} = -\frac{5}{4}. \quad (1.49)$$

[6] Carl Gustav Jacob Jacobi, 1804–1851, German/Prussian mathematician.

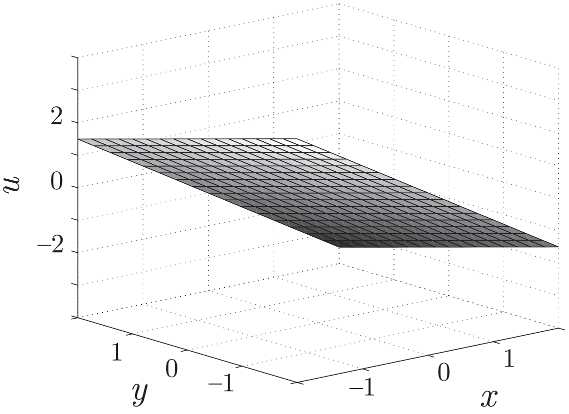


Figure 1.4. Planar surface, $u(x, y) = -5x/4 - y/2$, in the (x, y, u) volume formed by intersection of $x + y + u + v = 0$ and $2x - y + u - 3v = 0$.

Such an approach is generally easiest if $u(x, y)$ is explicitly available. Let us imagine that it is unavailable and show how we can obtain the same result with our more general approach. Equations (1.45) and (1.46) are rewritten as

$$f(x, y, u, v) = x + y + u + v = 0, \tag{1.50}$$

$$g(x, y, u, v) = 2x - y + u - 3v = 0. \tag{1.51}$$

Using the formula from Eq. (1.41) to solve for the desired derivative, we get

$$\frac{\partial u}{\partial x} = \frac{\begin{vmatrix} -\frac{\partial f}{\partial x} & \frac{\partial f}{\partial v} \\ -\frac{\partial g}{\partial x} & \frac{\partial g}{\partial v} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix}}. \tag{1.52}$$

Substituting, we get

$$\frac{\partial u}{\partial x} = \frac{\begin{vmatrix} -1 & 1 \\ -2 & -3 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 1 & -3 \end{vmatrix}} = \frac{3 - (-2)}{-3 - 1} = -\frac{5}{4}, \tag{1.53}$$

as expected. Here $J = -4$ for all real values of x and y (also written as $\forall\, x, y \in \mathbb{R}$), so we can always form $u(x, y)$ and $v(x, y)$.

Let us next consider a similar example, except that it is globally nonlinear.

EXAMPLE 1.5

If

$$x + y + u^6 + u + v = 0, \tag{1.54}$$

$$xy + uv = 1, \tag{1.55}$$

find $\partial u/\partial x$.

We again have four unknowns in two equations. In principle, we could solve for $u(x, y)$ and $v(x, y)$ and then determine all partial derivatives, such as the one desired. In practice, this is not always possible; for example, there is no general solution to sixth-order polynomial equations such as we have here (see Section A.1). The two hypersurfaces defined

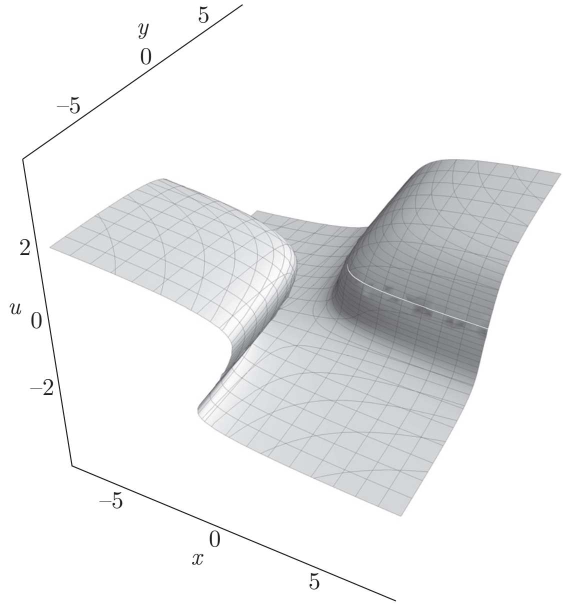


Figure 1.5. Nonplanar surface in the (x, y, u) volume formed by intersection of $x + y + u^6 + u + v = 0$ and $xy + uv = 1$.

by Eqs. (1.54) and (1.55) do in fact intersect to form a surface, $u(x, y)$. The surface can be obtained numerically and is plotted in Figure 1.5. For a given x and y, one can see that u may be nonunique.

To obtain the required partial derivative, the general method of this section is the most convenient. Equations (1.54) and (1.55) are rewritten as

$$f(x, y, u, v) = x + y + u^6 + u + v = 0, \tag{1.56}$$

$$g(x, y, u, v) = xy + uv - 1 = 0. \tag{1.57}$$

Using the formula from Eq. (1.41) to solve for the desired derivative, we get

$$\frac{\partial u}{\partial x} = \frac{\begin{vmatrix} -\frac{\partial f}{\partial x} & \frac{\partial f}{\partial v} \\ -\frac{\partial g}{\partial x} & \frac{\partial g}{\partial v} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix}}. \tag{1.58}$$

Substituting, we get

$$\frac{\partial u}{\partial x} = \frac{\begin{vmatrix} -1 & 1 \\ -y & u \end{vmatrix}}{\begin{vmatrix} 6u^5 + 1 & 1 \\ v & u \end{vmatrix}} = \frac{y - u}{u(6u^5 + 1) - v}. \tag{1.59}$$

Although Eq. (1.59) formally solves the problem at hand, we can perform some further analysis to study some of the features of the surface $u(x, y)$. Note if

$$v = 6u^6 + u \tag{1.60}$$

that the relevant Jacobian determinant is zero. At such points, determination of either $\partial u/\partial x$ or $\partial u/\partial y$ may be difficult or impossible; thus, for such points, it may not be possible to form $u(x, y)$. Following lengthy algebra, one can eliminate y and v from Eq. (1.59) to arrive at

$$\frac{\partial u}{\partial x} = \frac{1 + 2u^2 + u^7}{-1 + u^2 + 6u^7 - 2ux - 7u^6x - x^2}. \tag{1.61}$$

From Eq. (1.61), we see a potential singularity[7] in $\partial u/\partial x$ if the denominator is zero. Again, after some algebra, one finds a parametric representation of the curve for which $\partial u/\partial x$ may be singular:

$$x(s) = \frac{1}{2}\left(-2s - 7s^6 \pm \sqrt{-4 + 8s^2 + 52s^7 + 49s^{12}}\right), \tag{1.62}$$

$$y(s) = \frac{-1 - s^2 - s^7 - \frac{1}{2}s\left(-2s - 7s^6 \pm \sqrt{-4 + 8s^2 + 52s^7 + 49s^{12}}\right)}{s + \frac{1}{2}\left(2s + 7s^6 \mp \sqrt{-4 + 8s^2 + 52s^7 + 49s^{12}}\right)}, \tag{1.63}$$

$$u(s) = s. \tag{1.64}$$

Here s is a parameter, introduced for convenience. Certainly, on examination of Figure 1.5, we see there is a set of points for which $\partial u/\partial x$ is infinite.

We also note that $\partial u/\partial x = 0$ if the numerator of Eq. (1.61) is zero (assuming that the corresponding denominator is nonzero):

$$1 + 2u^2 + u^7 = 0. \tag{1.65}$$

This has seven roots, one of which is real, that being found by numerical solution to be $u = -1.21746$. Detailed algebra then reveals the parameterized curve for which $\partial u/\partial x = 0$ to be

$$x(s) = s, \tag{1.66}$$

$$y(s) = \frac{-1.21746 - s}{1 + 0.821383s}, \tag{1.67}$$

$$u(s) = -1.21746. \tag{1.68}$$

This is a curve in (x, y) space confined to a plane of constant u.

In summary, at points where the relevant Jacobian determinant $J = \partial(f, g)/\partial(u, v) \neq 0$ (which for this example includes nearly all of the (x, y) plane), given a local value of (x, y), we can use numerical algebra to find a corresponding u and v, which may be multivalued, and use the formula developed to find the local value of any required partial derivatives.

1.2 Inverse Function Theorem

Consider an explicit general mapping from x and y to u and v:

$$u = u(x, y), \tag{1.69}$$

$$v = v(x, y). \tag{1.70}$$

One can ask under what circumstances one can invert so that given u and v, one might find x and y. Relevant to this question, we have the following.

Inverse Function Theorem: If a mapping $u(x, y), v(x, y)$ is continuously differentiable and its Jacobian is invertible at a point (x_0, y_0), then u and v are uniquely invertible at that point and in its neighborhood.

The theorem can be extended to N functions of N variables.

[7] The word *singular* can have many connotations in mathematics, including the one here, which results from a potential division by zero. Other common usages are introduced later, e.g. in Sections 3.11 and 7.9.8.

EXAMPLE 1.6

Apply the inverse function theorem to analyze the unique invertibility of the mapping

$$u(x, y) = x + y, \tag{1.71}$$

$$v(x, y) = x - y. \tag{1.72}$$

The Jacobian $\mathbf{J}$ of the mapping is given by considering the total differentials, which in general matrix form are

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}}_{\mathbf{J}} \begin{pmatrix} dx \\ dy \end{pmatrix}. \tag{1.73}$$

For our problem, we find

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{J}} \begin{pmatrix} dx \\ dy \end{pmatrix}. \tag{1.74}$$

The derivative is invertible if $J = \det \mathbf{J} \neq 0$. Here we find $J = -2$. Because $J \neq 0$, the unique inverse always exists and is in fact

$$x = \frac{1}{2}(u + v), \tag{1.75}$$

$$y = \frac{1}{2}(u - v). \tag{1.76}$$

EXAMPLE 1.7

Altering the mapping of the previous example to

$$u = x + y, \tag{1.77}$$

$$v = 2x + 2y, \tag{1.78}$$

we find $J = 0$ everywhere, and the unique inverse mapping never exists. It is straightforward to see that u and v are not independent here as $u = v/2$ for all values of x and y. If u and v are given and $u \neq v/2$, it is impossible to find values for x and y. However, if $u = v/2$, not only can one find values for x and y but there are an infinite number of available solutions. Certainly, then, there is no unique solution for x and y. For example, if $(u, v) = (1, 2)$, we find $(x, y) = (1/2, 1/2)$ satisfies, as do an infinite number of other points, such as $(x, y) = (3/2, -1/2)$ or $(x, y) = (1, 0)$. In general $(x, y) = (1/2 + t, 1/2 - t)$ for $t \in (-\infty, \infty)$ will satisfy if $(u, v) = (1, 2)$. Geometrically, we can say that throughout (x, y) space, lines on which u is constant are parallel to lines on which v is constant. So if lines of constant u intersect lines of constant v, they do so in an infinite number of points. Systems of equations such as this are known as underconstrained and are considered in detail in Section 7.3.2.

EXAMPLE 1.8

Apply the inverse function theorem to analyze the unique invertibility of the mapping

$$u(x, y) = x + y^2, \tag{1.79}$$

$$v(x, y) = y + x^3. \tag{1.80}$$

The Jacobian $\mathbf{J}$ of the mapping is given by considering the total differentials, which in general matrix form are

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}}_{\mathbf{J}} \begin{pmatrix} dx \\ dy \end{pmatrix}. \tag{1.81}$$

For our problem, we find

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 2y \\ 3x^2 & 1 \end{pmatrix}}_{\mathbf{J}} \begin{pmatrix} dx \\ dy \end{pmatrix}. \tag{1.82}$$

The derivative is invertible if $J = \det \mathbf{J} \neq 0$. Here we find

$$J = 1 - 6x^2 y. \tag{1.83}$$

So if $y = 1/(6x^2)$, we have $J = 0$, and the unique inverse mapping does not exist for points in the neighborhood of this curve. As we see in the next section, if $J = 0$, we have that u is functionally dependent on v.

Consider the point $P : (x, y) = (1, 1/6)$, where we happen to have $J = 0$. At P, we have $(u, v) = (37/36, 7/6)$. At, and in the neighborhood of, P, we have the linearization

$$u - \frac{37}{36} = (x - 1) + \frac{1}{3}\left(y - \frac{1}{6}\right), \tag{1.84}$$

$$v - \frac{7}{6} = 3(x - 1) + \left(y - \frac{1}{6}\right). \tag{1.85}$$

Let us consider if there are other points than $P : (x, y) = (1, 1/6)$ that bring (u, v) into $(37/36, 7/6)$. Then we seek (x, y) such that

$$0 = (x - 1) + \frac{1}{3}\left(y - \frac{1}{6}\right), \tag{1.86}$$

$$0 = 3(x - 1) + \left(y - \frac{1}{6}\right). \tag{1.87}$$

This can be achieved for $(x, y) = (1 + t, 1/6 - 3t)$ where $t \in (-\infty, \infty)$, at least in the local linearization. Clearly, at P, there is no unique inverse function valid in the neighborhood of P. Once again, the nonunique nature of the solution in the neighborhood of P is a consequence of the linearized system being underconstrained and is considered in detail in Section 7.3.2.

If we multiply Eq. (1.85) by $-1/3$ and add it to Eq. (1.84), we eliminate both x and y and obtain the functional relation between u and v:

$$u = \frac{37}{36} + \frac{1}{3}\left(v - \frac{7}{6}\right). \tag{1.88}$$

This demonstrates how, if $J = 0$, we obtain a functional dependency between the otherwise independent u and v. At P, and at all points where $y = 1/6/x^2$, one finds that locally, lines of constant u are parallel to lines of constant v within (x, y) space.

1.3 Functional Dependence

Here we formalize notions regarding functional dependence. Let $u = u(x, y)$ and $v = v(x, y)$. If we can write $u = g(v)$ or $v = h(u)$, then u and v are said to be *functionally dependent*. If functional dependence between u and v exists, then we can consider $f(u,v)=0$. So, one can differentiate and expand $df=0$ to obtain the local linearization

$$\frac{\partial f}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial f}{\partial v}\frac{\partial v}{\partial x}=0, \tag{1.89}$$

$$\frac{\partial f}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial f}{\partial v}\frac{\partial v}{\partial y}=0. \tag{1.90}$$

In matrix form, this is

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{pmatrix}\begin{pmatrix} \frac{\partial f}{\partial u} \\ \frac{\partial f}{\partial v} \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1.91}$$

Because the right-hand side is zero, and we desire a nontrivial solution, Cramer's rule (see Eq. (7.95) or Section A.2), tells us that the determinant of the coefficient matrix must be zero for functional dependency, that is,

$$\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{vmatrix}=0. \tag{1.92}$$

Because it can be shown that $\det \mathbf{J}=\det \mathbf{J}^T=J$, this is equivalent to

$$J=\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}=\frac{\partial(u,v)}{\partial(x,y)}=0. \tag{1.93}$$

That is, the Jacobian determinant J must be zero for functional dependence. As an aside, we note that we employ the common notation of a superscript T, which signifies the transpose operation associated with exchanging nondiagonal terms across the diagonal of the matrix. It will be used from here on, but is first formally defined in Section 2.1.5.

EXAMPLE 1.9

Determine if

$$u=y+z, \tag{1.94}$$

$$v=x+2z^2, \tag{1.95}$$

$$w=x-4yz-2y^2 \tag{1.96}$$

are functionally dependent.

The determinant of the resulting coefficient matrix, by extension to three functions of three variables, is

$$\frac{\partial(u,v,w)}{\partial(x,y,z)}=\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{vmatrix}=\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} & \frac{\partial w}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} & \frac{\partial w}{\partial y} \\ \frac{\partial u}{\partial z} & \frac{\partial v}{\partial z} & \frac{\partial w}{\partial z} \end{vmatrix}, \tag{1.97}$$

$$=\begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & -4(y+z) \\ 1 & 4z & -4y \end{vmatrix}, \tag{1.98}$$

$$=(-1)(-4y-(-4)(y+z))+(1)(4z), \tag{1.99}$$

$$= 4y - 4y - 4z + 4z, \quad (1.100)$$

$$= 0. \quad (1.101)$$

So, u, v, w are functionally dependent. In fact, $w = v - 2u^2$, as can be verified by direct substitution.

EXAMPLE 1.10

Let

$$x + y + z = 0, \quad (1.102)$$

$$x^2 + y^2 + z^2 + 2xz = 1. \quad (1.103)$$

Can x and y be considered as functions of z?

If $x = x(z)$ and $y = y(z)$, then dx/dz and dy/dz must exist. If we take

$$f(x,y,z) = x + y + z = 0, \quad (1.104)$$

$$g(x,y,z) = x^2 + y^2 + z^2 + 2xz - 1 = 0, \quad (1.105)$$

$$df = \frac{\partial f}{\partial z}\,dz + \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = 0, \quad (1.106)$$

$$dg = \frac{\partial g}{\partial z}\,dz + \frac{\partial g}{\partial x}\,dx + \frac{\partial g}{\partial y}\,dy = 0, \quad (1.107)$$

$$\frac{\partial f}{\partial z} + \frac{\partial f}{\partial x}\frac{dx}{dz} + \frac{\partial f}{\partial y}\frac{dy}{dz} = 0, \quad (1.108)$$

$$\frac{\partial g}{\partial z} + \frac{\partial g}{\partial x}\frac{dx}{dz} + \frac{\partial g}{\partial y}\frac{dy}{dz} = 0, \quad (1.109)$$

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{dx}{dz} \\ \frac{dy}{dz} \end{pmatrix} = \begin{pmatrix} -\frac{\partial f}{\partial z} \\ -\frac{\partial g}{\partial z} \end{pmatrix}, \quad (1.110)$$

then the solution[8] $(dx/dz, dy/dz)^T$ can be obtained by Cramer's rule:

$$\frac{dx}{dz} = \frac{\begin{vmatrix} -\frac{\partial f}{\partial z} & \frac{\partial f}{\partial y} \\ -\frac{\partial g}{\partial z} & \frac{\partial g}{\partial y} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix}} = \frac{\begin{vmatrix} -1 & 1 \\ -(2z+2x) & 2y \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 2x+2z & 2y \end{vmatrix}} = \frac{-2y + 2z + 2x}{2y - 2x - 2z} = -1, \quad (1.111)$$

$$\frac{dy}{dz} = \frac{\begin{vmatrix} \frac{\partial f}{\partial x} & -\frac{\partial f}{\partial z} \\ \frac{\partial g}{\partial x} & -\frac{\partial g}{\partial z} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix}} = \frac{\begin{vmatrix} 1 & -1 \\ 2x+2z & -(2z+2x) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 2x+2z & 2y \end{vmatrix}} = \frac{0}{2y - 2x - 2z}, \quad (1.112)$$

In Eq. (1.111), the numerator and denominator cancel; there is no special condition defined by the Jacobian determinant of the denominator being zero. In Eq. (1.112),

[8] Note that the transpose operation can also be applied to one-dimensional matrices, sometimes described as vectors. This has the effect of converting a so-called row vector into a column vector, and vice versa. This will be considered in detail in Chapter 2.

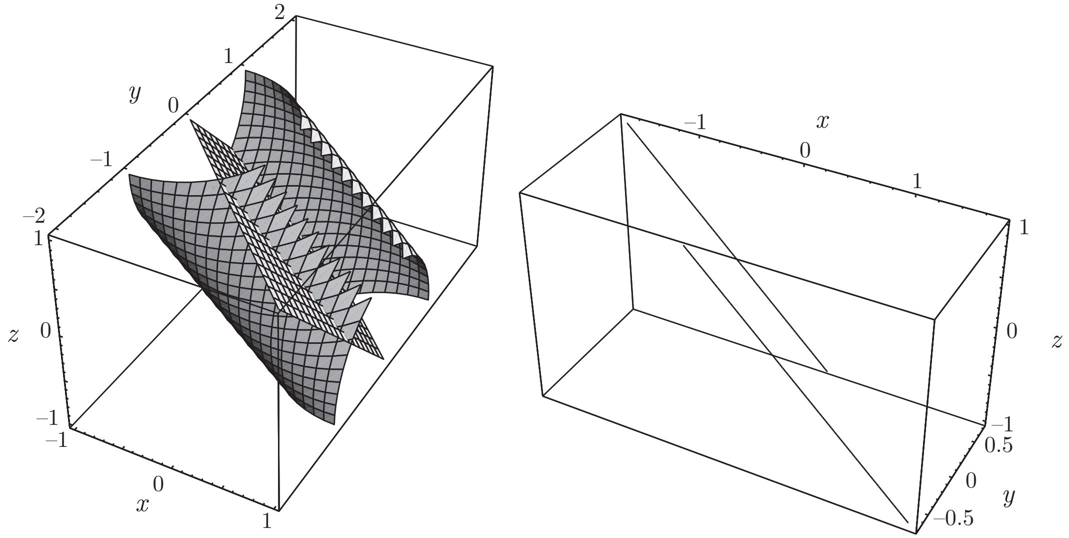


Figure 1.6. Surfaces of $x + y + z = 0$ and $x^2 + y^2 + z^2 + 2xz = 1$, and their loci of intersection.

$dy/dz = 0$ if $y - x - z \neq 0$. If $y - x - z = 0$, Eq. (1.112) is indeterminate for purposes of defining dy/dz, and we cannot be certain that $y(z)$ exists.

Now it is easily shown by algebraic manipulations (which for more general functions are not possible) that $x(z)$ and $y(z)$ do exist and are given by

$$x(z) = -z \pm \frac{\sqrt{2}}{2}, \tag{1.113}$$

$$y(z) = \mp\frac{\sqrt{2}}{2}. \tag{1.114}$$

This forms two distinct lines in (x, y, z) space. On these lines, $J = 2y - 2x - 2z = \mp 2\sqrt{2}$, which is never zero.

The two original functions and their loci of intersection are plotted in Figure 1.6. It is seen that the surface represented by the linear function Eq. (1.102) is a plane, and that represented by the quadratic function Eq. (1.103) is an open cylindrical tube. Note that planes and cylinders may or may not intersect. If they intersect, it is most likely that the intersection will be a closed arc. However, when the plane is parallel to the axis of the cylinder, if an intersection exists, it will be in two parallel lines; such is the case here.

Let us see how slightly altering the equation for the plane removes the degeneracy. Take now

$$5x + y + z = 0, \tag{1.115}$$

$$x^2 + y^2 + z^2 + 2xz = 1. \tag{1.116}$$

Can x and y be considered as functions of z? If $x = x(z)$ and $y = y(z)$, then dx/dz and dy/dz must exist. If we take

$$f(x, y, z) = 5x + y + z = 0, \tag{1.117}$$

$$g(x, y, z) = x^2 + y^2 + z^2 + 2xz - 1 = 0, \tag{1.118}$$

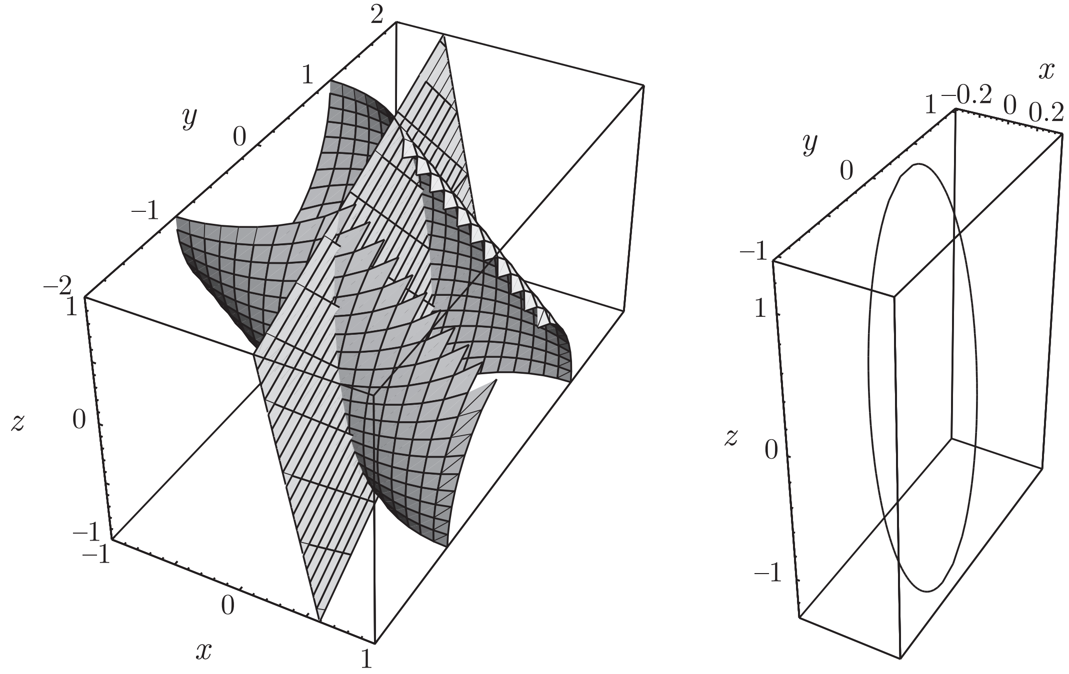


Figure 1.7. Surfaces of $5x + y + z = 0$ and $x^2 + y^2 + z^2 + 2xz = 1$, and their loci of intersection.

then the solution $(dx/dz, dy/dz)^T$ is found as before:

$$\frac{dx}{dz} = \frac{\begin{vmatrix} -\dfrac{\partial f}{\partial z} & \dfrac{\partial f}{\partial y} \\ -\dfrac{\partial g}{\partial z} & \dfrac{\partial g}{\partial y} \end{vmatrix}}{\begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{vmatrix}} = \frac{\begin{vmatrix} -1 & 1 \\ -(2z+2x) & 2y \end{vmatrix}}{\begin{vmatrix} 5 & 1 \\ 2x+2z & 2y \end{vmatrix}} = \frac{-2y+2z+2x}{10y-2x-2z}, \tag{1.119}$$

$$\frac{dy}{dz} = \frac{\begin{vmatrix} \dfrac{\partial f}{\partial x} & -\dfrac{\partial f}{\partial z} \\ \dfrac{\partial g}{\partial x} & -\dfrac{\partial g}{\partial z} \end{vmatrix}}{\begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{vmatrix}} = \frac{\begin{vmatrix} 5 & -1 \\ 2x+2z & -(2z+2x) \end{vmatrix}}{\begin{vmatrix} 5 & 1 \\ 2x+2z & 2y \end{vmatrix}} = \frac{-8x-8z}{10y-2x-2z}. \tag{1.120}$$

The two original functions and their loci of intersection are plotted in Figure 1.7.

Straightforward algebra in this case shows that an explicit dependency exists:

$$x(z) = \frac{-6z \pm \sqrt{2}\sqrt{13-8z^2}}{26}, \tag{1.121}$$

$$y(z) = \frac{-4z \mp 5\sqrt{2}\sqrt{13-8z^2}}{26}. \tag{1.122}$$

These curves represent the projection of the curve of intersection on the x, z and y, z planes, respectively. In both cases, the projections are ellipses. Considered in three dimensions, the two equations can be thought of as the parameterization of a curve, with z as the parameter. In (x, y, z) space, that curve is also an ellipse. With considerable effort, not shown here, using the methods to be discussed in Sections 7.7.2 and 7.14, we can describe the ellipse by

$$\frac{2}{27}\left(11+\sqrt{13}\right)\left(\frac{\left(3\sqrt{3}-2\sqrt{13}\right)\hat{y}}{5\sqrt{1+\frac{1}{25}\left(2\sqrt{13}-3\sqrt{3}\right)^2}}+\frac{\hat{z}}{\sqrt{1+\frac{1}{25}\left(2\sqrt{13}-3\sqrt{3}\right)^2}}\right)^2$$
$$+\frac{2}{27}\left(11-\sqrt{13}\right)\left(\frac{\left(-3\sqrt{3}-2\sqrt{13}\right)\hat{y}}{5\sqrt{1+\frac{1}{25}\left(3\sqrt{3}+2\sqrt{13}\right)^2}}-\frac{\hat{z}}{\sqrt{1+\frac{1}{25}\left(3\sqrt{3}+2\sqrt{13}\right)^2}}\right)^2=1, \tag{1.123}$$

where $\hat{y}$ and $\hat{z}$ are given by

$$\hat{y}=\frac{x}{3\sqrt{3}}+\frac{1}{18}\left(9-5\sqrt{3}\right)y+\frac{1}{18}\left(-9-5\sqrt{3}\right)z, \tag{1.124}$$

$$\hat{z}=\frac{x}{3\sqrt{3}}+\frac{1}{18}\left(-9-5\sqrt{3}\right)y+\frac{1}{18}\left(9-5\sqrt{3}\right)z. \tag{1.125}$$

This form can be achieved by first finding a linear transformation from (x, y, z) to $(\hat{x}, \hat{y}, \hat{z})$ that represents the plane $5x + y + z = 0$ as $\hat{x} = 0$. We have employed Eqs. (1.124, 1.125), coupled with

$$\hat{x}=\frac{5x}{3\sqrt{3}}+\frac{y}{3\sqrt{3}}+\frac{z}{3\sqrt{3}}, \tag{1.126}$$

to effect the variable change. This choice is nonunique, and there are others that achieve the same end with a simpler form. Our choice was selected so that, among other things, a particular triple (x, y, z) will have the same magnitude as its transform $(\hat{x}, \hat{y}, \hat{z})$. Then one studies the cylinder in this transformed space, imposes $\hat{x} = 0$ to effect the intersection with the plane, and uses algebraic manipulations to put it into the form of an ellipse.

1.4 Leibniz Rule

Often functions are expressed in terms of integrals. For example,

$$y(x)=\int_{a(x)}^{b(x)} f(x,t)\,dt. \tag{1.127}$$

Here t is a dummy variable of integration. Although x and t appear in Eq. (1.127), because t is removed by the integration process, it is strictly speaking not multivariable in the same sense as previous sections. However, it is necessary to consider expressions of this type in the upcoming Section 1.5.2, and it is easily extended to truly multivariable scenarios; see Section 2.8.5. We often need to take derivatives of functions in the form expressed in Eq. (1.127), and the *Leibniz*[9] *rule* tells us how to do so. For such functions, the Leibniz rule gives us

$$\frac{dy(x)}{dx}=f(x,b(x))\frac{db(x)}{dx}-f(x,a(x))\frac{da(x)}{dx}+\int_{a(x)}^{b(x)}\frac{\partial f(x,t)}{\partial x}\,dt. \tag{1.128}$$

[9] Gottfried Wilhelm von Leibniz, 1646–1716, German mathematician and philosopher; co-inventor with Sir Isaac Newton, 1643–1727, of the calculus.

EXAMPLE 1.11

If

$$y(x) = y(x_0) + \int_{x_0}^{x} f(t)\,dt, \tag{1.129}$$

use the Leibniz rule to find dy/dx.

Here $f(x,t)$ simply reduces to $f(t)$; we also have $b(x) = x$ and $a(x) = x_0$. Applying the Leibniz rule, we find

$$\frac{dy(x)}{dx} = \underbrace{\frac{d}{dx}(y(x_0))}_{=0} + f(x)\underbrace{\frac{d}{dx}(x)}_{=1} - f(x_0)\underbrace{\frac{d}{dx}(x_0)}_{=0} + \int_{x_0}^{x} \underbrace{\frac{\partial}{\partial x}(f(t))}_{=0}\,dt, \tag{1.130}$$

$$= f(x). \tag{1.131}$$

The integral expression naturally includes the initial condition that if $x = x_0$, $y = y(x_0)$. This needs to be expressed separately for the differential version of the equation.

EXAMPLE 1.12

Find dy/dx if

$$y(x) = \int_{x}^{x^2} (x+1)t^2\,dt. \tag{1.132}$$

Using the Leibniz rule, we get

$$\frac{dy(x)}{dx} = ((x+1)x^4)(2x) - ((x+1)x^2)(1) + \int_{x}^{x^2} t^2\,dt, \tag{1.133}$$

$$= 2x^6 + 2x^5 - x^3 - x^2 + \left(\frac{t^3}{3}\right)\Bigg|_{x}^{x^2}, \tag{1.134}$$

$$= 2x^6 + 2x^5 - x^3 - x^2 + \frac{x^6}{3} - \frac{x^3}{3}, \tag{1.135}$$

$$= \frac{7x^6}{3} + 2x^5 - \frac{4x^3}{3} - x^2. \tag{1.136}$$

In this case, it is possible to integrate explicitly to achieve the same result:

$$y(x) = (x+1)\int_{x}^{x^2} t^2\,dt, \tag{1.137}$$

$$= (x+1)\left(\frac{t^3}{3}\right)\Bigg|_{x}^{x^2}, \tag{1.138}$$

$$= (x+1)\left(\frac{x^6}{3} - \frac{x^3}{3}\right), \tag{1.139}$$

$$= \frac{x^7}{3} + \frac{x^6}{3} - \frac{x^4}{3} - \frac{x^3}{3}, \tag{1.140}$$

$$\frac{dy(x)}{dx} = \frac{7x^6}{3} + 2x^5 - \frac{4x^3}{3} - x^2. \tag{1.141}$$

So the two methods give identical results.

As an aside, let us take up a common issue of notation, related to our discussion of the Leibniz rule as well as many other topics in the mathematics of engineering systems. Let us imagine we are presented with the differential equation

$$\frac{dy}{dx} = f(x), \quad y(x_0) = y_0. \tag{1.142}$$

One might be tempted to propose the solution to be $y(x) = y_0 + \int_{x_0}^{x} f(x)\, dx$. However, this is in error! The statement is in fact incoherent. Certainly, if $x = x_0$, it predicts the correct value of $y(x_0) = y_0$ because an integral over a domain of width zero has no value. However, if we attempt to evaluate if $x = x_1 \neq x_0$, we see a problem: $y(x_1) = y_0 + \int_{x_0}^{x_1} f(x_1)\, dx_1$ does not make sense in many ways. First, $f(x_1)$ is a constant and could be brought outside of the integral. Second, because x_1 is a constant, $dx_1 = 0$, yielding an incorrect interpretation of $y(x_1) = y_0$. Because the variable of integration must change through the domain of integration, it is important that it have a different nomenclature than the upper limit.

Let us proceed more systematically through the use of dummy variables. First, introduce the dummy variable t into Eq. (1.142) by trading it for x:

$$\frac{dy}{dt} = f(t). \tag{1.143}$$

Now, apply the operator $\int_{x_0}^{x} (\cdot)\, dt$ to both sides:

$$\int_{x_0}^{x} \frac{dy}{dt}\, dt = \int_{x_0}^{x} f(t)\, dt. \tag{1.144}$$

The fundamental theorem of calculus[10] allows evaluation of the integral on the left side to yield

$$y(x) - y(x_0) = \int_{x_0}^{x} f(t)\, dt, \tag{1.145}$$

$$y(x) = y_0 + \int_{x_0}^{x} f(t)\, dt. \tag{1.146}$$

Application of the Leibniz rule to this solution yields the correct $dy/dx = f(x)$.

1.5 Optimization

The basic strategies of optimization have motivated calculus from its inception. Here we consider a variety of approaches for multivariable optimization.[11] We first consider how to select specific points that optimize nonlinear functions of many variables in the absence of external constraints. This is followed by a discussion of how to select specific functions to optimize integrals that take those functions as

[10] This well-known theorem has two parts: (1) if $f(x)$ is continuous for $x \in [a, b]$ and the function F can be shown to be the indefinite integral of f in the domain $x \in [a, b]$, then $\int_a^b f(x)\, dx = F(b) - F(a)$, and (2) if on an open interval $F(x) = \int_a^x f(t)\, dt$, then $dF/dx = f(x)$ for a within the interval.

[11] Though the general term *optimization* includes other aspects, we are concerned here only with extremization that is associated with zero derivatives and functions that are not extreme at boundaries.

inputs. The section is completed by a discussion of how to optimize in the presence of external constraints.

1.5.1 Unconstrained Optimization

Consider the real function $f(x)$, where $x \in [a, b]$. Extrema are at $x = x_m$, where $f'(x_m) = 0$, if $x_m \in [a, b]$. It is a local minimum, a local maximum, or an inflection point, according to whether $f''(x_m)$ is positive, negative, or zero, respectively.

Now consider a function of two variables $f(x, y)$, with $x \in [a, b]$, $y \in [c, d]$. A necessary condition for an extremum is

$$\frac{\partial f}{\partial x}(x_m, y_m) = \frac{\partial f}{\partial y}(x_m, y_m) = 0, \tag{1.147}$$

where $x_m \in [a, b]$, $y_m \in [c, d]$. Next, we find the Hessian[12] matrix:

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}. \tag{1.148}$$

We use $\mathbf{H}$ and its elements to determine the character of any local extremum. It can be shown that

- f is a maximum if $\partial^2 f/\partial x^2 < 0$, $\partial^2 f/\partial y^2 < 0$, and $\partial^2 f/\partial x \partial y < \sqrt{(\partial^2 f/\partial x^2)(\partial^2 f/\partial y^2)}$,
- f is a minimum if $\partial^2 f/\partial x^2 > 0$, $\partial^2 f/\partial y^2 > 0$, and $\partial^2 f/\partial x \partial y < \sqrt{(\partial^2 f/\partial x^2)(\partial^2 f/\partial y^2)}$,
- f is a saddle otherwise, as long as $\det \mathbf{H} \neq 0$,
- if $\det \mathbf{H} = 0$, higher order terms need to be considered.

The first two conditions for maximum and minimum require that terms on the diagonal of $\mathbf{H}$ dominate those on the off-diagonal with diagonal terms further required to be of the same sign. For higher dimensional systems, one can show that if all the eigenvalues of $\mathbf{H}$ are negative, f is maximized, and if all the eigenvalues of $\mathbf{H}$ are positive, f is minimized.

One can begin to understand this by considering a Taylor series expansion of $f(x, y)$. Taking $\mathbf{x} = (x, y)^T$, $d\mathbf{x} = (dx, dy)^T$, and $\nabla f = (\partial f/\partial x, \partial f/\partial y)^T$, multivariable Taylor series expansion gives

$$f(\mathbf{x} + d\mathbf{x}) = f(\mathbf{x}) + d\mathbf{x}^T \cdot \underbrace{\nabla f}_{=0} + d\mathbf{x}^T \cdot \mathbf{H} \cdot d\mathbf{x} + \cdots. \tag{1.149}$$

At an extremum, $\nabla f = 0$, so

$$f(\mathbf{x} + d\mathbf{x}) = f(\mathbf{x}) + d\mathbf{x}^T \cdot \mathbf{H} \cdot d\mathbf{x} + \cdots. \tag{1.150}$$

Later (see Section 6.7.1 and Section 7.2.5), we see that, by virtue of the definition of the term *positive definite*, if the Hessian $\mathbf{H}$ is positive definite, then $\forall\ d\mathbf{x} \neq \mathbf{0}$, $d\mathbf{x}^T \cdot \mathbf{H} \cdot d\mathbf{x} > 0$, which corresponds to a minimum. If $\mathbf{H}$ is positive definite within a domain, we say that the surface described by f is *convex*, a topic further considered in Section 6.3. For negative definite $\mathbf{H}$, we have a maximum.

[12] Ludwig Otto Hesse, 1811–1874, German mathematician.

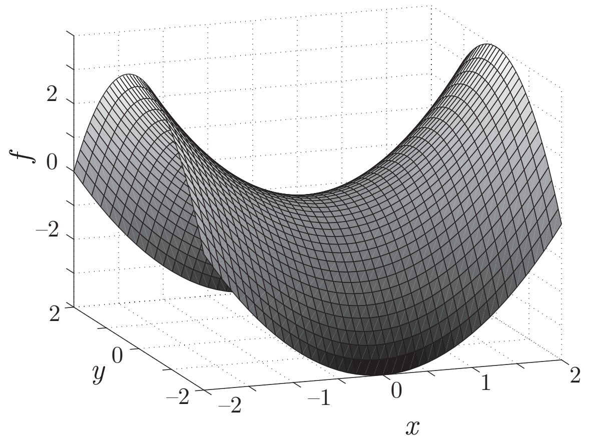


Figure 1.8. Surface of $f = x^2 - y^2$ showing saddle character at the critical point $(x, y) = (0, 0)$.

EXAMPLE 1.13

Find the extrema of

$$f = x^2 - y^2. \tag{1.151}$$

Equating partial derivatives with respect to x and to y to zero, we get

$$\frac{\partial f}{\partial x} = 2x = 0, \tag{1.152}$$

$$\frac{\partial f}{\partial y} = -2y = 0. \tag{1.153}$$

This gives $x = 0, y = 0$. For this f, we find

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}, \tag{1.154}$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}. \tag{1.155}$$

Because (1) $\det \mathbf{H} = -4 \neq 0$ and (2) $\partial^2 f/\partial x^2$ and $\partial^2 f/\partial y^2$ have different signs, the equilibrium is a saddle point. The saddle shape of the surface is visible in the plot of Figure 1.8.

1.5.2 Calculus of Variations

The problem of the *calculus of variations* is to find the function $y(x)$, with $x \in [x_1, x_2]$, and boundary conditions $y(x_1) = y_1, y(x_2) = y_2$, such that

$$I = \int_{x_1}^{x_2} f(x, y, y')\, dx \tag{1.156}$$

is an extremum. For efficiency, we adopt here the common shorthand notation of *prime* denoting differentiation with respect to x, that is, $dy/dx = y'$. In Eq. (1.156), we encounter the operation of mapping a function $y(x)$ into a scalar I, which can be expressed as $I = \mathbf{F}y(x)$. The operator $\mathbf{F}$ that performs this task is known as a *functional*.

If $y(x)$ is the desired solution, let $Y(x) = y(x) + \epsilon h(x)$, where $h(x_1) = h(x_2) = 0$. Thus, $Y(x)$ also satisfies the boundary conditions; also, $Y'(x) = y'(x) + \epsilon h'(x)$. We can write

$$I(\epsilon) = \int_{x_1}^{x_2} f(x, Y, Y')\, dx. \tag{1.157}$$

Taking $dI/d\epsilon$, we get

$$\frac{dI}{d\epsilon} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial x} \underbrace{\frac{\partial x}{\partial \epsilon}}_{0} + \frac{\partial f}{\partial Y} \underbrace{\frac{\partial Y}{\partial \epsilon}}_{h(x)} + \frac{\partial f}{\partial Y'} \underbrace{\frac{\partial Y'}{\partial \epsilon}}_{h'(x)} \right) dx. \tag{1.158}$$

Evaluating, we find

$$\frac{dI}{d\epsilon} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial x} 0 + \frac{\partial f}{\partial Y} h(x) + \frac{\partial f}{\partial Y'} h'(x) \right) dx. \tag{1.159}$$

Because I is an extremum at $\epsilon = 0$, we have $dI/d\epsilon = 0$ for $\epsilon = 0$. This gives

$$0 = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial Y} h(x) + \frac{\partial f}{\partial Y'} h'(x) \right) \bigg|_{\epsilon=0} dx. \tag{1.160}$$

When $\epsilon = 0$, we have $Y = y, Y' = y'$, so

$$0 = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} h(x) + \frac{\partial f}{\partial y'} h'(x) \right) dx. \tag{1.161}$$

We can look at the second term in this integral. From integration by parts,[13] we get

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'} h'(x)\, dx = \int_{x_1}^{x_2} \frac{\partial f}{\partial y'} \frac{dh}{dx}\, dx = \int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\, dh, \tag{1.162}$$

$$= \underbrace{\frac{\partial f}{\partial y'} h(x) \bigg|_{x_1}^{x_2}}_{=0} - \int_{x_1}^{x_2} \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) h(x)\, dx, \tag{1.163}$$

$$= - \int_{x_1}^{x_2} \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) h(x)\, dx. \tag{1.164}$$

The first term in Eq. (1.163) is zero because of our conditions on $h(x_1)$ and $h(x_2)$. Thus, substituting Eq. (1.164) into the original equation, Eq. (1.161), we find

$$\int_{x_1}^{x_2} \underbrace{\left(\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right)}_{0} h(x)\, dx = 0. \tag{1.165}$$

[13] Integration by parts tells us $\int u(x) v'(x)\, dx = u(x)v(x) - \int u'(x) v(x)\, dx$, or in short, $\int u\, dv = uv - \int v\, du$.

The equality holds for all $h(x)$, so that we must have

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0. \tag{1.166}$$

Equation (1.166) is called the *Euler*[14]*-Lagrange*[15] *equation*; sometimes it is simply called Euler's equation.

While this is, in general, the preferred form of the Euler-Lagrange equation, its explicit dependency on the two end conditions is better displayed by considering a slightly different form. By expanding the term that is differentiated with respect to x, we obtain

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}(x,y,y')\right) = \frac{\partial^2 f}{\partial y'\partial x}\underbrace{\frac{dx}{dx}}_{=1} + \frac{\partial^2 f}{\partial y'\partial y}\underbrace{\frac{dy}{dx}}_{y'} + \frac{\partial^2 f}{\partial y'\partial y'}\underbrace{\frac{dy'}{dx}}_{y''}, \tag{1.167}$$

$$= \frac{\partial^2 f}{\partial y'\partial x} + \frac{\partial^2 f}{\partial y'\partial y}y' + \frac{\partial^2 f}{\partial y'\partial y'}y''. \tag{1.168}$$

Thus, the Euler-Lagrange equation, Eq. (1.166), after slight rearrangement, becomes

$$\frac{\partial^2 f}{\partial y'\partial y'}y'' + \frac{\partial^2 f}{\partial y'\partial y}y' + \frac{\partial^2 f}{\partial y'\partial x} - \frac{\partial f}{\partial y} = 0, \tag{1.169}$$

$$f_{y'y'}y'' + f_{y'y}y' + (f_{y'x} - f_y) = 0. \tag{1.170}$$

This is a second-order differential equation for $f_{y'y'} \neq 0$ and, in general, nonlinear. If $f_{y'y'}$ is always nonzero, the problem is said to be *regular*. If $f_{y'y'} = 0$, the equation is no longer second order, and the problem is said to be *singular* at such points. This connotation of the word "singular" is used whenever the highest order term, here y'', is multiplied by zero. Satisfaction of two boundary conditions becomes problematic for equations less than second order.

There are several special cases of the function f.

- $f = f(x,y)$:
  The Euler-Lagrange equation is

  $$\frac{\partial f}{\partial y} = 0, \tag{1.171}$$

  which is easily solved:

  $$f(x,y) = A(x), \tag{1.172}$$

  which, knowing f, is then solved for $y(x)$.
- $f = f(x,y')$:
  The Euler-Lagrange equation is

  $$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0, \tag{1.173}$$

[14] Leonhard Euler, 1707–1783, Swiss mathematician.
[15] Joseph-Louis Lagrange, 1736–1813, Italian-born French mathematician.

which yields

$$\frac{\partial f}{\partial y'} = A, \tag{1.174}$$

$$f(x, y') = Ay' + B(x). \tag{1.175}$$

Again, knowing f, the equation is solved for y' and then integrated to find $y(x)$.

- $f = f(y, y')$:
  The Euler-Lagrange equation is

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}(y, y')\right) = 0, \tag{1.176}$$

$$\frac{\partial f}{\partial y} - \left(\frac{\partial^2 f}{\partial y \partial y'}\frac{dy}{dx} + \frac{\partial^2 f}{\partial y' \partial y'}\frac{dy'}{dx}\right) = 0, \tag{1.177}$$

$$\frac{\partial f}{\partial y} - \frac{\partial^2 f}{\partial y \partial y'}\frac{dy}{dx} - \frac{\partial^2 f}{\partial y' \partial y'}\frac{d^2 y}{dx^2} = 0. \tag{1.178}$$

Multiply by y' to arrive at

$$y'\left(\frac{\partial f}{\partial y} - \frac{\partial^2 f}{\partial y \partial y'}\frac{dy}{dx} - \frac{\partial^2 f}{\partial y' \partial y'}\frac{d^2 y}{dx^2}\right) = 0. \tag{1.179}$$

Add and subtract $(\partial f/\partial y')y''$ to find

$$y'\left(\frac{\partial f}{\partial y} - \frac{\partial^2 f}{\partial y \partial y'}\frac{dy}{dx} - \frac{\partial^2 f}{\partial y' \partial y'}\frac{d^2 y}{dx^2}\right) + \frac{\partial f}{\partial y'}y'' - \frac{\partial f}{\partial y'}y'' = 0. \tag{1.180}$$

Regroup to get

$$\underbrace{\frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial y'}y''}_{=df/dx} - \underbrace{\left(y'\left(\frac{\partial^2 f}{\partial y \partial y'}\frac{dy}{dx} + \frac{\partial^2 f}{\partial y' \partial y'}\frac{d^2 y}{dx^2}\right) + \frac{\partial f}{\partial y'}y''\right)}_{=d/dx(y'\partial f/\partial y')} = 0. \tag{1.181}$$

Regroup again to show

$$\frac{d}{dx}\left(f - y'\frac{\partial f}{\partial y'}\right) = 0, \tag{1.182}$$

which can be integrated. Thus,

$$f(y, y') - y'\frac{\partial f}{\partial y'} = K, \tag{1.183}$$

where K is an arbitrary constant. What remains is a first-order ordinary differential equation that can be solved. Another integration constant arises. This second constant along with K are determined by the two endpoint conditions.

EXAMPLE 1.14

Find the curve of minimum length between the points (x_1, y_1) and (x_2, y_2).

If $y(x)$ is the curve, then $y(x_1) = y_1$ and $y(x_2) = y_2$. In the most common geometries, the length of the curve can be shown[16] to be

$$\ell = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}\, dx. \tag{1.184}$$

[16] If $d\ell^2 = dx^2 + dy^2$, then $d\ell = \sqrt{1 + (dy/dx)^2}\, dx = \sqrt{1 + (y')^2}\, dx$, which can be integrated to find ℓ.

So our f reduces to $f(y') = \sqrt{1+(y')^2}$. The Euler-Lagrange equation is

$$\frac{d}{dx}\left(\frac{y'}{\sqrt{1+(y')^2}}\right) = 0, \tag{1.185}$$

which can be integrated to give

$$\frac{y'}{\sqrt{1+(y')^2}} = K. \tag{1.186}$$

Solving for y', we get

$$y' = \sqrt{\frac{K^2}{1-K^2}} \equiv A, \tag{1.187}$$

from which

$$y = Ax + B. \tag{1.188}$$

The constants A and B are obtained from the boundary conditions $y(x_1) = y_1$ and $y(x_2) = y_2$. *The shortest distance between two points is a straight line.*

EXAMPLE 1.15

Find the curve through the points (x_1, y_1) and (x_2, y_2), such that the surface area of the body of revolution found by rotating the curve around the x-axis is a minimum.

We wish to minimize the integral that yields the surface area of a body of revolution:

$$I = \int_{x_1}^{x_2} y\sqrt{1+(y')^2}\, dx. \tag{1.189}$$

Here f reduces to $f(y, y') = y\sqrt{1+(y')^2}$. So the Euler-Lagrange equation reduces to

$$f(y, y') - y'\frac{\partial f}{\partial y'} = A, \tag{1.190}$$

$$y\sqrt{1+y'^2} - y'y\frac{y'}{\sqrt{1+y'^2}} = A, \tag{1.191}$$

$$y(1+y'^2) - yy'^2 = A\sqrt{1+y'^2}, \tag{1.192}$$

$$y = A\sqrt{1+y'^2}, \tag{1.193}$$

$$y' = \sqrt{\left(\frac{y}{A}\right)^2 - 1}, \tag{1.194}$$

$$y(x) = A\cosh\frac{x-B}{A}. \tag{1.195}$$

This is a catenary. The constants A and B are determined from the boundary conditions $y(x_1) = y_1$ and $y(x_2) = y_2$. In general, this requires a trial-and-error solution of simultaneous algebraic equations. If $(x_1, y_1) = (-1, 3.08616)$ and $(x_2, y_2) = (2, 2.25525)$, one finds that solution of the resulting algebraic equations gives $A = 2, B = 1$. For these conditions, the curve $y(x)$ along with the resulting body of revolution of minimum surface area are plotted in Figure 1.9.

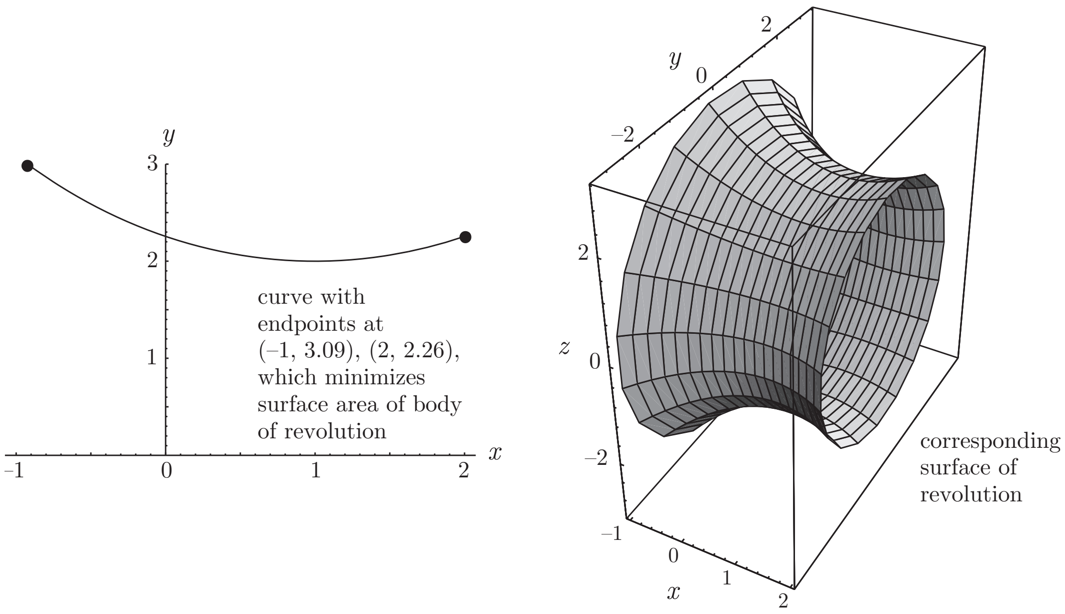


Figure 1.9. Body of revolution of minimum surface area for $(x_1, y_1) = (-1, 3.08616)$ and $(x_2, y_2) = (2, 2.25525)$.

EXAMPLE 1.16

Examine the consequences of the extremization of

$$I = \int_{t_1}^{t_2} \left(\frac{1}{2} m\dot{y}^2 - mgy \right) dt, \tag{1.196}$$

where m and g are constants.

Here we have made a trivial change of independent variables from x to t, and simultaneously changed the derivative notation from $'$ to $\dot{}$. The reason for these small changes is that this is a problem that has special significance in mechanics if we consider m to be a mass, g to be a constant gravitational acceleration, y to be position, and t to be time. Within mechanics, a physical principle in which some I is extremized is known as a *variational principle*. And within mechanics, the variational principle we are considering here is known as the principle of least action. Thus, the action that will be least is I. One also often describes $m\dot{y}^2/2$ as the kinetic energy T, mgy as the potential energy V, and their difference as the Lagrangian L. That is, $L = T - V = m\dot{y}^2/2 - mgy$.

So we are considering the equivalent of a limiting case of Eq. (1.156) if it takes the form

$$I = \int_{t_1}^{t_2} f(y, \dot{y})\, dt, \tag{1.197}$$

where

$$f(y, \dot{y}) = \frac{1}{2} m\dot{y}^2 - mgy. \tag{1.198}$$

If we assert that I is extremized, that is, we assert the variational principle holds, we can immediately state that many forms of the appropriate Euler-Lagrange equation hold.

One important form is the equivalent of Eq. (1.166), which reduces to

$$\frac{\partial f}{\partial y} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0, \tag{1.199}$$

$$-mg - \frac{d}{dt}(m\dot{y}) = 0, \tag{1.200}$$

$$m\frac{d^2y}{dt^2} = -mg. \tag{1.201}$$

This is simply Newton's celebrated second law of motion for a body accelerating under the influence of a constant gravitational force. One might say for this case that Newton's second law is one and the same as the variational principle of least action.

Another form arises from the appropriate version of Eq. (1.183), selecting for convenient physical interpretation $K = -C$:

$$f(y, \dot{y}) - \dot{y}\frac{\partial f}{\partial \dot{y}} = -C, \tag{1.202}$$

$$\frac{1}{2}m\dot{y}^2 - mgy - \dot{y}m\dot{y} = -C, \tag{1.203}$$

$$\frac{1}{2}m\dot{y}^2 + mgy = C. \tag{1.204}$$

This well-known equation is an expression of one type of energy conservation principle, valid in mechanics when no forces of dissipation are present. Here we have the kinetic, $m\dot{y}^2/2$, and potential, mgy, energies combining to form a constant, C.

1.5.3 Constrained Optimization: Lagrange Multipliers

Suppose we have to determine the extremum of $f(x_1, x_2, \ldots, x_M)$ subject to the N constraints

$$g_n(x_1, x_2, \ldots, x_M) = 0, \qquad n = 1, 2, \ldots, N. \tag{1.205}$$

Define

$$f^* = f - \lambda_1 g_1 - \lambda_2 g_2 - \cdots - \lambda_N g_N, \tag{1.206}$$

where the λ_n $(n = 1, 2, \ldots, N)$ are unknown constants called *Lagrange multipliers*. To get the extremum of f^*, we equate to zero its derivative with respect to $x_1, x_2, \ldots, x_M$. Thus, we have the complete set

$$\frac{\partial f^*}{\partial x_m} = 0, \; m = 1, \ldots, M, \tag{1.207}$$

$$g_n = 0, \; n = 1, \ldots, N, \tag{1.208}$$

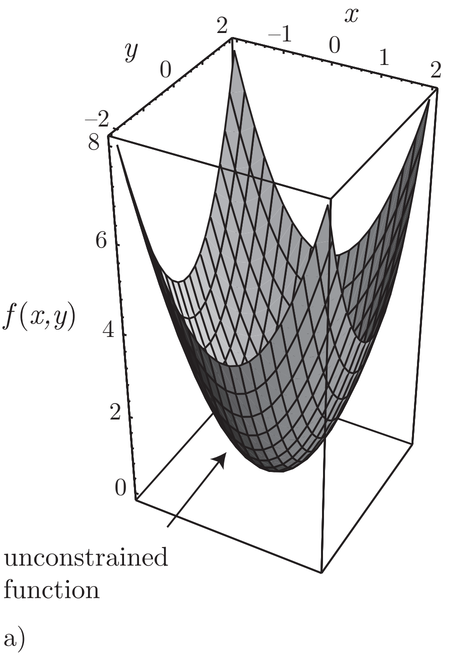
which are $(M + N)$ equations that can be solved for x_m $(m = 1, 2, \ldots, M)$ and λ_n $(n = 1, 2, \ldots, N)$.

EXAMPLE 1.17

Extremize $f = x^2 + y^2$ subject to the constraint $g = 5x^2 - 6xy + 5y^2 - 8 = 0$.

Here we have $N = 1$ constraint for a function in $M = 2$ variables. Let

$$f^* = x^2 + y^2 - \lambda(5x^2 - 6xy + 5y^2 - 8), \tag{1.209}$$



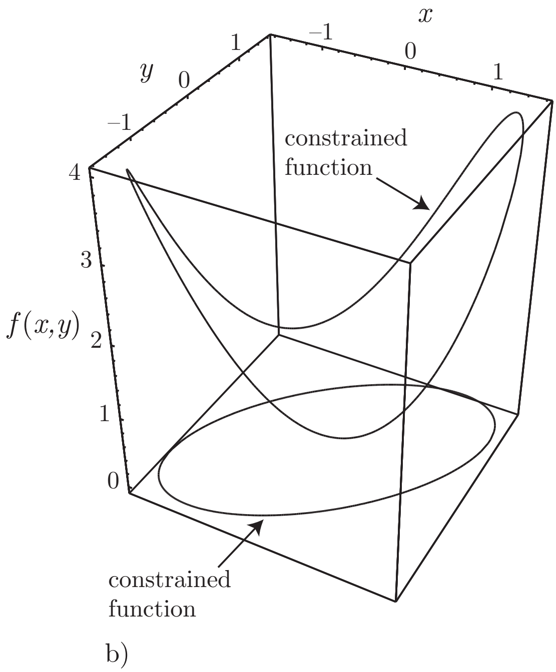


Figure 1.10. (a) Unconstrained function $f(x, y)$ and (b) constrained function and constraint function (image of constrained function.)

from which

$$\frac{\partial f^*}{\partial x} = 2x - 10\lambda x + 6\lambda y = 0, \tag{1.210}$$

$$\frac{\partial f^*}{\partial y} = 2y + 6\lambda x - 10\lambda y = 0, \tag{1.211}$$

$$g = 5x^2 - 6xy + 5y^2 - 8 = 0. \tag{1.212}$$

From Eq. (1.210),

$$\lambda = \frac{2x}{10x - 6y}, \tag{1.213}$$

which, if substituted into Eq. (1.211), gives

$$x = \pm y. \tag{1.214}$$

Equation (1.214), when solved in conjunction with Eq. (1.212), gives the extrema to be at $(x, y) = (\sqrt{2}, \sqrt{2})$, $(-\sqrt{2}, -\sqrt{2})$, $(1/\sqrt{2}, -1/\sqrt{2})$, $(-1/\sqrt{2}, 1/\sqrt{2})$. The first two sets give $f = 4$ (maximum) and the last two $f = 1$ (minimum). The function to be maximized along with the constraint function and its image are plotted in Figure 1.10.

A similar technique can be used for the extremization of a functional with constraints. We wish to find the function $y(x)$, with $x \in [x_1, x_2]$, and $y(x_1) = y_1, y(x_2) = y_2$, such that the integral

$$I = \int_{x_1}^{x_2} f(x, y, y')\, dx \tag{1.215}$$

is an extremum and satisfies the constraint

$$g = 0. \tag{1.216}$$

Define

$$I^* = I - \lambda g, \tag{1.217}$$

and continue as before.

EXAMPLE 1.18

Extremize I, where

$$I = \int_0^a y\sqrt{1 + (y')^2}\, dx, \tag{1.218}$$

with $y(0) = y(a) = 0$, subject to the constraint

$$\int_0^a \sqrt{1 + (y')^2}\, dx = \ell. \tag{1.219}$$

That is, find the maximum surface area of a body of revolution whose generating curve has a constant length.

Let

$$g = \int_0^a \sqrt{1 + (y')^2}\, dx - \ell = 0. \tag{1.220}$$

Then let

$$I^* = I - \lambda g = \int_0^a y\sqrt{1 + (y')^2}\, dx - \lambda \int_0^a \sqrt{1 + (y')^2}\, dx + \lambda\ell, \tag{1.221}$$

$$= \int_0^a (y - \lambda)\sqrt{1 + (y')^2}\, dx + \lambda\ell, \tag{1.222}$$

$$= \int_0^a \left((y - \lambda)\sqrt{1 + (y')^2} + \frac{\lambda\ell}{a} \right) dx. \tag{1.223}$$

With $f^* = (y - \lambda)\sqrt{1 + (y')^2} + \lambda\ell/a$, we have the Euler-Lagrange equation

$$\frac{\partial f^*}{\partial y} - \frac{d}{dx}\left(\frac{\partial f^*}{\partial y'}\right) = 0. \tag{1.224}$$

Integrating from an earlier developed relationship, Eq. (1.183), when $f = f(y, y')$, and absorbing $\lambda\ell/a$ into a constant A, we have

$$(y - \lambda)\sqrt{1 + (y')^2} - y'(y - \lambda)\frac{y'}{\sqrt{1 + (y')^2}} = A, \tag{1.225}$$

from which

$$(y - \lambda)(1 + (y')^2) - (y')^2(y - \lambda) = A\sqrt{1 + (y')^2}, \tag{1.226}$$

$$(y - \lambda)\left(1 + (y')^2 - (y')^2\right) = A\sqrt{1 + (y')^2}, \tag{1.227}$$

$$y - \lambda = A\sqrt{1 + (y')^2}, \tag{1.228}$$

$$y' = \sqrt{\left(\frac{y - \lambda}{A}\right)^2 - 1}, \tag{1.229}$$

$$y = \lambda + A\cosh\frac{x - B}{A}. \tag{1.230}$$

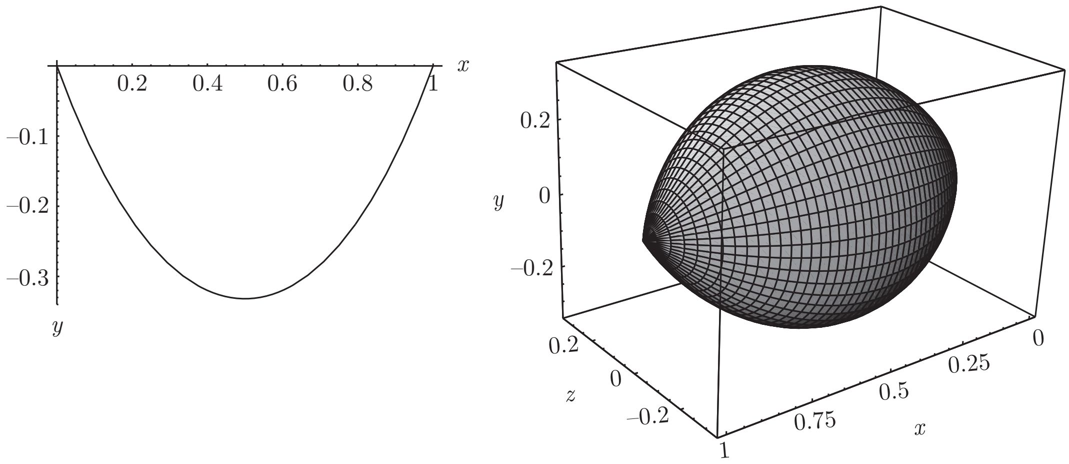


Figure 1.11. Curve of length $\ell = 5/4$ with $y(0) = y(1) = 0$ whose surface area of corresponding body of revolution (also shown) is maximum.

Here A, B, λ have to be numerically determined from the three conditions $y(0) = y(a) = 0, g = 0$. If we take $a = 1, \ell = 5/4$, we find that $A = 0.422752, B = 1/2, \lambda = -0.754549$. For these values, the curve of interest, along with the surface of revolution, is plotted in Figure 1.11.

1.6 Non-Cartesian Coordinate Transformations

Many problems are formulated in ordinary three-dimensional Cartesian[17] space, typically represented by the three variables (x, y, z). However, many of these problems, especially those involving curved geometrical bodies, are more efficiently posed in a non-Cartesian, curvilinear coordinate system. To facilitate analysis involving such geometries, one needs techniques to transform from one coordinate system to another. The transformations that effect this are usually expressed in terms of algebraic relationships between a set of variables, thus rendering the methods of multivariable calculus to be relevant.

In this section, we focus on general nonlinear transformations involving three spatial variables. The analysis can be adapted for an arbitrary number of variables. This general framework is the appropriate one for a variety of tasks in engineering and physics, though it introduces complexities. Coordinate transformations that are Cartesian, the subject of the upcoming Chapter 2, are simpler and often useful, but lack the flexibility afforded by those of this section.

For this section, we utilize the so-called *Einstein*[18] *index notation*. Thus, we take untransformed Cartesian coordinates, most commonly known as (x, y, z), to be represented instead by (ξ^1, ξ^2, ξ^3); that is to say $x \to \xi^1$, $y \to \xi^2$ and $z \to \xi^3$. *Here the superscript is an index and does not represent a power of* ξ. We denote a point by ξ^i, where $i = 1, 2, 3$. A nearby point will be located at $\xi^i + d\xi^i$, where $d\xi^i$ is a differential distance. Because the space is Cartesian, we have the usual Euclidean[19]

[17] René Descartes, 1596–1650, French mathematician and philosopher.
[18] Albert Einstein, 1879–1955, German-American physicist and mathematician.
[19] Euclid of Alexandria, c. 325–265 B.C., Greek geometer.

formula for the distance between the point and the nearby point from Pythagoras'[20] theorem for a differential arc length ds:

$$(ds)^2 = \left(d\xi^1\right)^2 + \left(d\xi^2\right)^2 + \left(d\xi^3\right)^2, \tag{1.231}$$

$$= \sum_{i=1}^{3} d\xi^i \, d\xi^i \equiv d\xi^i \, d\xi^i. \tag{1.232}$$

Here we have adopted Einstein's summation convention that *if an index appears twice, a summation from 1 to 3 is understood*. Though it makes little difference here, to strictly adhere to the conventions of the Einstein notation, which require a balance of sub- and superscripts, we should more formally take

$$(ds)^2 = d\xi^j \, \delta_{ji} \, d\xi^i = d\xi_i \, d\xi^i, \tag{1.233}$$

where δ_{ji} is the *Kronecker*[21] *delta*,

$$\delta_{ji} = \delta^{ji} = \delta^i_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{1.234}$$

In matrix form, the Kronecker delta is simply the identity matrix $\mathbf{I}$, that is,

$$\delta_{ji} = \delta^{ji} = \delta^i_j = \mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.235}$$

In matrix form, we can also represent Eq. (1.233) as

$$(ds)^2 = \begin{pmatrix} d\xi_1 & d\xi_2 & d\xi_3 \end{pmatrix} \begin{pmatrix} d\xi^1 \\ d\xi^2 \\ d\xi^3 \end{pmatrix} = d\boldsymbol{\xi}^T \cdot d\boldsymbol{\xi}. \tag{1.236}$$

We associate subscripted triples of scalars with row vectors and superscripted triples of scalars with column vectors. In the final expression of Eq. (1.236), we introduce the familiar bold-faced typography to represent vectors; this is known as *Gibbs*[22] *notation.*[23]

Now let us consider a point P whose representation in Cartesian coordinates is (ξ^1, ξ^2, ξ^3) and transform those coordinates so that P is represented in a more convenient (x^1, x^2, x^3) space. This transformation is achieved by defining the following functional dependencies:

$$x^1 = x^1(\xi^1, \xi^2, \xi^3), \tag{1.237}$$

$$x^2 = x^2(\xi^1, \xi^2, \xi^3), \tag{1.238}$$

$$x^3 = x^3(\xi^1, \xi^2, \xi^3). \tag{1.239}$$

We are thus considering arbitrary nonlinear algebraic functions that take ξ^i into x^j, $i, j = 1, 2, 3$. In this discussion, we make the common presumption that the entity P is invariant and that it has different representations in different coordinate systems.

[20] Pythagoras of Samos, c. 570–490 B.C., Ionian Greek mathematician, philosopher, and mystic.
[21] Leopold Kronecker, 1823–1891, German-Prussian mathematician.
[22] Josiah Willard Gibbs, 1839–1903, American mechanical engineer and mathematician.
[23] Here, for the Gibbs notation, we choose to explicitly indicate that the operation requires the first term be transposed to form a row vector. Many notation schemes assume this to be an intrinsic part of the operation and would simply state $(ds)^2 = d\boldsymbol{\xi} \cdot d\boldsymbol{\xi}$.

Thus, the coordinate axes change, but the location of P does not. This is known as an *alias* transformation. This contrasts another common approach in which a point is represented in an original space, and after application of a transformation, it is again represented in the original space in an altered state. This is known as an *alibi* transformation. The alias approach transforms the axes; the alibi approach transforms the elements of the space. Throughout this chapter, as well as in Chapter 2, we are especially concerned with alias transformations, which allow us to represent the same entity in different ways. In other chapters, especially Chapter 7, we focus more on alibi transformations.

Taking derivatives can tell us whether the inverse mapping exists. The derivatives of x^1, x^2, and x^3 are

$$dx^1 = \frac{\partial x^1}{\partial \xi^1}\, d\xi^1 + \frac{\partial x^1}{\partial \xi^2}\, d\xi^2 + \frac{\partial x^1}{\partial \xi^3}\, d\xi^3 = \frac{\partial x^1}{\partial \xi^j}\, d\xi^j, \tag{1.240}$$

$$dx^2 = \frac{\partial x^2}{\partial \xi^1}\, d\xi^1 + \frac{\partial x^2}{\partial \xi^2}\, d\xi^2 + \frac{\partial x^2}{\partial \xi^3}\, d\xi^3 = \frac{\partial x^2}{\partial \xi^j}\, d\xi^j, \tag{1.241}$$

$$dx^3 = \frac{\partial x^3}{\partial \xi^1}\, d\xi^1 + \frac{\partial x^3}{\partial \xi^2}\, d\xi^2 + \frac{\partial x^3}{\partial \xi^3}\, d\xi^3 = \frac{\partial x^3}{\partial \xi^j}\, d\xi^j. \tag{1.242}$$

In matrix form, these reduce to

$$\begin{pmatrix} dx^1 \\ dx^2 \\ dx^3 \end{pmatrix} = \begin{pmatrix} \frac{\partial x^1}{\partial \xi^1} & \frac{\partial x^1}{\partial \xi^2} & \frac{\partial x^1}{\partial \xi^3} \\ \frac{\partial x^2}{\partial \xi^1} & \frac{\partial x^2}{\partial \xi^2} & \frac{\partial x^2}{\partial \xi^3} \\ \frac{\partial x^3}{\partial \xi^1} & \frac{\partial x^3}{\partial \xi^2} & \frac{\partial x^3}{\partial \xi^3} \end{pmatrix} \begin{pmatrix} d\xi^1 \\ d\xi^2 \\ d\xi^3 \end{pmatrix}. \tag{1.243}$$

In Einstein notation, this greatly simplifies to

$$dx^i = \frac{\partial x^i}{\partial \xi^j}\, d\xi^j. \tag{1.244}$$

For the inverse to exist, we must have a nonzero Jacobian determinant for the transformation, that is,

$$\frac{\partial(x^1, x^2, x^3)}{\partial(\xi^1, \xi^2, \xi^3)} \neq 0. \tag{1.245}$$

As long as Eq. (1.245) is satisfied, the inverse transformation exists:

$$\xi^1 = \xi^1(x^1, x^2, x^3), \tag{1.246}$$

$$\xi^2 = \xi^2(x^1, x^2, x^3), \tag{1.247}$$

$$\xi^3 = \xi^3(x^1, x^2, x^3). \tag{1.248}$$

Likewise, then,

$$d\xi^i = \frac{\partial \xi^i}{\partial x^j}\, dx^j. \tag{1.249}$$