Lecture Notes on
The Mechanics of Elastic Solids

Volume 1: A Brief Review of Some Mathematical Preliminaries

Version 1.0

Rohan Abeyaratne
Quentin Berg Professor of Mechanics
Department of Mechanical Engineering
MIT

http://web.mit.edu/abeyaratne/lecture_notes.html

December 2, 2006

Electronic Publication

Rohan Abeyaratne
Quentin Berg Professor of Mechanics
Department of Mechanical Engineering
77 Massachusetts Institute of Technology
Cambridge, MA 02139-4307, USA

Abeyaratne, Rohan, 1952-
Lecture Notes on The Mechanics of Elastic Solids. Volume 1: A Brief Review of Some Mathematical Preliminaries / Rohan Abeyaratne – 1st Edition – Cambridge, MA:

ISBN-13: 978-0-9791865-0-9
ISBN-10: 0-9791865-0-1

QC

Please send corrections, suggestions and comments to *abeyaratne.vol.1@gmail.com*

Updated June 25 2007

Dedicated with admiration and affection
to Matt Murphy and the miracle of science,
for the gift of renaissance.

PREFACE

The Department of Mechanical Engineering at MIT offers a series of graduate level subjects on the Mechanics of Solids and Structures which include:

2.071: Mechanics of Solid Materials,
2.072: Mechanics of Continuous Media,
2.074: Solid Mechanics: Elasticity,
2.073: Solid Mechanics: Plasticity and Inelastic Deformation,
2.075: Advanced Mechanical Behavior of Materials,
2.080: Structural Mechanics,
2.094: Finite Element Analysis of Solids and Fluids,
2.095: Molecular Modeling and Simulation for Mechanics, and
2.099: Computational Mechanics of Materials.

Over the years, I have had the opportunity to regularly teach the second and third of these subjects, 2.072 and 2.074 (formerly known as 2.083), and the current three volumes are comprised of the lecture notes I developed for them. The first draft of these notes was produced in 1987 and they have been corrected, refined and expanded on every following occasion that I taught these classes. The material in the current presentation is still meant to be a set of lecture notes, not a text book. It has been organized as follows:

Volume I: A Brief Review of Some Mathematical Preliminaries

Volume II: Continuum Mechanics

Volume III: Elasticity

My appreciation for mechanics was nucleated by Professors Douglas Amarasekara and Munidasa Ranaweera of the (then) University of Ceylon, and was subsequently shaped and grew substantially under the influence of Professors James K. Knowles and Eli Sternberg of the California Institute of Technology. I have been most fortunate to have had the opportunity to apprentice under these inspiring and distinctive scholars. I would especially like to acknowledge a great many illuminating and stimulating interactions with my mentor, colleague and friend Jim Knowles, whose influence on me cannot be overstated.

I am also indebted to the many MIT students who have given me enormous fulfillment and joy to be part of their education.

My understanding of elasticity as well as these notes have also benefitted greatly from many useful conversations with Kaushik Bhattacharya, Janet Blume, Eliot Fried, Morton E.

Gurtin, Richard D. James, Stelios Kyriakides, David M. Parks, Phoebus Rosakis, Stewart Silling and Nicolas Triantafyllidis, which I gratefully acknowledge.

Volume I of these notes provides a collection of essential definitions, results, and illustrative examples, designed to review those aspects of mathematics that will be encountered in the subsequent volumes. It is most certainly *not* meant to be a source for learning these topics for the first time. The treatment is concise, selective and limited in scope. For example, Linear Algebra is a far richer subject than the treatment here, which is limited to real 3-dimensional Euclidean vector spaces.

The topics covered in Volumes II and III are largely those one would expect to see covered in such a set of lecture notes. Personal taste has led me to include a few special (but still well-known) topics. Examples of this include sections on the statistical mechanical theory of polymer chains and the lattice theory of crystalline solids in the discussion of constitutive theory in Volume II; and sections on the so-called Eshelby problem and the effective behavior of two-phase materials in Volume III.

There are a number of Worked Examples at the end of each chapter which are an essential part of the notes. Many of these examples either provide, more details, or a proof, of a result that had been quoted previously in the text; or it illustrates a general concept; or it establishes a result that will be used subsequently (possibly in a later volume).

The content of these notes are entirely classical, in the best sense of the word, and none of the material here is original. I have drawn on a number of sources over the years as I prepared my lectures. I cannot recall every source I have used but certainly they include those listed at the end of each chapter. In a more general sense the broad approach and philosophy taken has been influenced by:

Volume I: A Brief Review of Some Mathematical Preliminaries

I.M. Gelfand and S.V. Fomin, *Calculus of Variations*, Prentice Hall, 1963.

J.K. Knowles, *Linear Vector Spaces and Cartesian Tensors*, Oxford University Press, New York, 1997.

Volume II: Continuum Mechanics

P. Chadwick, *Continuum Mechanics: Concise Theory and Problems*, Dover,1999.

J.L. Ericksen, *Introduction to the Thermodynamics of Solids*, Chapman and Hall, 1991.

M.E. Gurtin, *An Introduction to Continuum Mechanics*, Academic Press, 1981.

J. K. Knowles and E. Sternberg, *(Unpublished) Lecture Notes for AM136: Finite Elasticity*, California Institute of Technology, Pasadena, CA 1978.

C. Truesdell and W. Noll, The nonlinear field theories of mechanics, in *Handbüch der Physik*, Edited by S. Flügge, Volume III/3, Springer, 1965.

Volume IIII: Elasticity

M.E. Gurtin, The linear theory of elasticity, in *Mechanics of Solids - Volume II*, edited by C. Truesdell, Springer-Verlag, 1984.

J. K. Knowles, *(Unpublished) Lecture Notes for AM135: Elasticity*, California Institute of Technology, Pasadena, CA, 1976.

A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, Dover, 1944.

S. P. Timoshenko and J.N. Goodier, *Theory of Elasticity*, McGraw-Hill, 1987.

The following notation will be used consistently in Volume I: Greek letters will denote real numbers; lowercase boldface Latin letters will denote vectors; and uppercase boldface Latin letters will denote linear transformations. Thus, for example, $\alpha, \beta, \gamma ...$ will denote scalars (real numbers); $\mathbf{a}, \mathbf{b}, \mathbf{c}, ...$ will denote vectors; and $\mathbf{A}, \mathbf{B}, \mathbf{C}, ...$ will denote linear transformations. In particular, "**o**" will denote the null vector while "**0**" will denote the null linear transformation. As much as possible this notation will also be used in Volumes II and III though there will be some lapses (for reasons of tradition).

Contents

Chapter 1

Matrix Algebra and Indicial Notation

Notation:

$\{a\}$	.....	$m \times 1$ matrix, i.e. a column matrix with m rows and one column
a_i	.....	element in row-i of the column matrix $\{a\}$
$[A]$	.....	$m \times n$ matrix
A_{ij}	.....	element in row-i, column-j of the matrix $[A]$

1.1 Matrix algebra

Even though more general matrices can be considered, for our purposes it is sufficient to consider a matrix to be a rectangular array of real numbers that obeys certain rules of addition and multiplication. A $m \times n$ matrix $[A]$ has m rows and n columns:

$$[A] \;=\; \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix} \;; \tag{1.1}$$

A_{ij} denotes the element located in the *ith* row and *jth* column. The column matrix

$$\{x\} \;=\; \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{pmatrix} \tag{1.2}$$

has m rows and one column; The row matrix

$$\{y\} \;=\; \{y_1, y_2, \ldots, y_n\} \tag{1.3}$$

has one row and n columns. If all the elements of a matrix are zero it is said to be a *null matrix* and is denoted by $[0]$ or $\{0\}$ as the case may be.

Two $m\times n$ matrices $[A]$ and $[B]$ are said to be *equal* if and only if all of their corresponding elements are equal:

$$A_{ij} = B_{ij}, \qquad i = 1, 2, \ldots m, \quad j = 1, 2, \ldots, n. \tag{1.4}$$

If $[A]$ and $[B]$ are both $m \times n$ matrices, their *sum* is the $m \times n$ matrix $[C]$ denoted by $[C] = [A] + [B]$ whose elements are

$$C_{ij} = A_{ij} + B_{ij}, \qquad i = 1, 2, \ldots m, \quad j = 1, 2, \ldots, n. \tag{1.5}$$

If $[A]$ is a $p \times q$ matrix and $[B]$ is a $q \times r$ matrix, their *product* is the $p \times r$ matrix $[C]$ with elements

$$C_{ij} = \sum_{k=1}^{q} A_{ik}B_{kj}, \qquad i = 1, 2, \ldots p, \quad j = 1, 2, \ldots, q; \tag{1.6}$$

one writes $[C] = [A][B]$. In general $[A][B] \neq [B][A]$; therefore rather than referring to $[A][B]$ as the product of $[A]$ and $[B]$ we should more precisely refer to $[A][B]$ as $[A]$ postmultiplied by $[B]$; or $[B]$ premultiplied by $[A]$. It is worth noting that if two matrices $[A]$ and $[B]$ obey the equation $[A][B] = [0]$ this does not necessarily mean that either $[A]$ or $[B]$ has to be the null matrix $[0]$. Similarly if three matrices $[A], [B]$ and $[C]$ obey $[A][B] = [A][C]$ this does not necessarily mean that $[B] = [C]$ (even if $[A] \neq [0]$.) The *product by a scalar* α of a $m \times n$ matrix $[A]$ is the $m \times n$ matrix $[B]$ with components

$$B_{ij} = \alpha A_{ij}, \qquad i = 1, 2, \ldots m, \quad j = 1, 2, \ldots, n; \tag{1.7}$$

one writes $[B] = \alpha[A]$.

Note that a $m_1 \times n_1$ matrix $[A_1]$ can be postmultiplied by a $m_2 \times n_2$ matrix $[A_2]$ if and only if $n_1 = m_2$. In particular, consider a $m \times n$ matrix $[A]$ and a $n \times 1$ (column) matrix $\{x\}$. Then we can postmultiply $[A]$ by $\{x\}$ to get the $m \times 1$ column matrix $[A]\{x\}$; but we cannot premultiply $[A]$ by $\{x\}$ (unless m=1), i.e. $\{x\}[A]$ does not exist is general.

The *transpose* of the $m \times n$ matrix $[A]$ is the $n \times m$ matrix $[B]$ where

$$B_{ij} = A_{ji} \qquad \text{for each } i = 1, 2, \ldots n, \text{ and } j = 1, 2, \ldots, m. \tag{1.8}$$

Usually one denotes the matrix $[B]$ by $[A]^T$. One can verify that

$$[A+B]^T = [A]^T + [B]^T, \qquad [AB]^T = [B]^T[A]^T. \tag{1.9}$$

The transpose of a column matrix is a row matrix; and vice versa. Suppose that $[A]$ is a $m \times n$ matrix and that $\{x\}$ is a $m \times 1$ (column) matrix. Then we can premultiply $[A]$ by $\{x\}^T$, i.e. $\{x\}^T[A]$ exists (and is a $1 \times n$ row matrix). For any $n \times 1$ column matrix $\{x\}$ note that

$$\{x\}^T\{x\} = \{x\}\{x\}^T = x_1^2 + x_2^2 \ldots + x_n^2 = \sum_{i=1}^{n} x_i^2. \tag{1.10}$$

A $n \times n$ matrix $[A]$ is called a *square matrix*; the diagonal elements of this matrix are the A_{ii}'s. A square matrix $[A]$ is said to be *symmetrical* if

$$A_{ij} = A_{ji} \qquad \text{for each } i, j = 1, 2, \ldots n; \tag{1.11}$$

skew-symmetrical if

$$A_{ij} = -A_{ji} \qquad \text{for each } i, j = 1, 2, \ldots n. \tag{1.12}$$

Thus for a symmetric matrix $[A]$ we have $[A]^T = [A]$; for a skew-symmetric matrix $[A]$ we have $[A]^T = -[A]$. Observe that each diagonal element of a skew-symmetric matrix must be zero.

If the off-diagonal elements of a square matrix are all zero, i.e. $A_{ij} = 0$ for each $i, j = 1, 2, \ldots n, i \neq j$, the matrix is said to be *diagonal*. If every diagonal element of a diagonal matrix is 1 the matrix is called a *unit matrix* and is usually denoted by $[I]$.

Suppose that $[A]$ is a $n \times n$ square matrix and that $\{x\}$ is a $n \times 1$ (column) matrix. Then we can postmultiply $[A]$ by $\{x\}$ to get a $n \times 1$ column matrix $[A]\{x\}$, and premultiply the resulting matrix by $\{x\}^T$ to get a 1×1 square matrix, effectively just a scalar, $\{x\}^T[A]\{x\}$. Note that

$$\{x\}^T[A]\{x\} \; = \; \sum_{i=1}^{n} \sum_{j=1}^{n} \; A_{ij}x_ix_j. \tag{1.13}$$

This is referred to as the quadratic form associated with $[A]$. In the special case of a diagonal matrix $[A]$

$$\{x\}^T[A]\{x\} \; = \; A_{11}x_1^2 + A_{22}x_1^2 + \ldots + A_{nn}x_n^2. \tag{1.14}$$

The *trace* of a square matrix is the sum of the diagonal elements of that matrix and is denoted by trace$[A]$:

$$\text{trace}[A] = \sum_{i=1}^{n} \; A_{ii}. \tag{1.15}$$

One can show that

$$\text{trace}([A][B]) = \text{trace}([B][A]). \tag{1.16}$$

Let det$[A]$ denote the *determinant* of a square matrix. Then for a 2×2 matrix

$$\det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = A_{11}A_{22} - A_{12}A_{21}, \tag{1.17}$$

and for a 3×3 matrix

$$\det \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = A_{11} \det \begin{pmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix} - A_{12} \det \begin{pmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{pmatrix} + A_{13} \det \begin{pmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix}. \tag{1.18}$$

The determinant of a $n \times n$ matrix is defined recursively in a similar manner. One can show that

$$\det([A][B]) = (\det[A])\ (\det[B]). \tag{1.19}$$

Note that trace$[A]$ and det$[A]$ are both scalar-valued functions of the matrix $[A]$.

Consider a square matrix $[A]$. For each $i = 1, 2, \ldots, n$, a row matrix $\{a\}_i$ can be created by assembling the elements in the *ith* row of $[A]$: $\{a\}_i = \{A_{i1}, A_{i2}, A_{i3}, \ldots, A_{in}\}$. If the only scalars α_i for which

$$\alpha_1\{a\}_1 + \alpha_2\{a\}_2 + \alpha_3\{a\}_3 + \ldots \alpha_n\{a\}_n = \{0\} \tag{1.20}$$

are $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$, the rows of $[A]$ are said to be linearly independent. If at least one of the α's is non-zero, they are said to be linearly dependent, and then at least one row of $[A]$ can be expressed as a linear combination of the other rows.

Consider a square matrix $[A]$ and suppose that its rows are linearly independent. Then the matrix is said to be *non-singular* and there exists a matrix $[B]$, usually denoted by $[B] = [A]^{-1}$ and called the *inverse* of [A], for which $[B][A] = [A][B] = [I]$. For $[A]$ to be non-singular it is necessary and sufficient that $\det[A] \neq 0$. If the rows of $[A]$ are linearly dependent, the matrix is singular and an inverse matrix does not exist.

Consider a $n \times n$ square matrix $[A]$. First consider the $(n-1) \times (n-1)$ matrix obtained by eliminating the *ith* row and *jth* column of $[A]$; then consider the determinant of that second matrix; and finally consider the product of that determinant with $(-1)^{i+j}$. The number thus obtained is called the cofactor of A_{ij}. If $[B]$ is the inverse of $[A]$, $[B] = [A]^{-1}$, then

$$B_{ij} = \frac{\text{cofactor of } A_{ji}}{\det[A]} \tag{1.21}$$

If the transpose and inverse of a matrix coincide, i.e. if

$$[A]^{-1} = [A]^T, \tag{1.22}$$

then the matrix is said to be *orthogonal*. Note that for an orthogonal matrix $[A]$, one has $[A][A]^T = [A]^T[A] = [I]$ and that $\det[A] = \pm 1$.

1.2 Indicial notation

Consider a $n \times n$ square matrix $[A]$ and two $n \times 1$ column matrices $\{x\}$ and $\{b\}$. Let A_{ij} denote the element of $[A]$ in its i^{th} row and j^{th} column, and let x_i and b_i denote the elements in the i^{th} row of $\{x\}$ and $\{b\}$ respectively. Now consider the matrix equation $[A]\{x\} = \{b\}$:

$$\begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_n \end{pmatrix}. \tag{1.23}$$

Carrying out the matrix multiplication, this is equivalent to the system of linear algebraic equations

$$\left.\begin{array}{llllll} A_{11}x_1 & +A_{12}x_2 & +\ldots & +A_{1n}x_n & = & b_1, \\ A_{21}x_1 & +A_{22}x_2 & +\ldots & +A_{2n}x_n & = & b_2, \\ \ldots & +\ldots & +\ldots & +\ldots & =\ldots & \\ A_{n1}x_1 & +A_{n2}x_2 & +\ldots & +A_{nn}x_n & = & b_n. \end{array}\right\} \tag{1.24}$$

This system of equations can be written more compactly as

$$A_{i1}x_1 + A_{i2}x_2 + \ldots A_{in}x_n = b_i \quad \text{with } i \text{ taking each value in the range } 1, 2, \ldots n; \tag{1.25}$$

or even more compactly by omitting the statement "with i taking each value in the range $1, 2, \ldots, n$", and simply writing

$$A_{i1}x_1 + A_{i2}x_2 + \ldots + A_{in}x_n = b_i \tag{1.26}$$

with the understanding that (1.26) *holds for each value of the subscript i in the range $i = 1, 2, \ldots n$.* This understanding is referred to as the *range convention.* The subscript i is called a *free subscript* because it is free to take on each value in its range. From here on, we shall always use the range convention unless explicitly stated otherwise.

Observe that

$$A_{j1}x_1 + A_{j2}x_2 + \ldots + A_{jn}x_n = b_j \tag{1.27}$$

is identical to (1.26); this is because j is a free subscript in (1.27) and so (1.27) is required to hold "for all $j = 1, 2, \ldots, n$" and this leads back to (1.24). This illustrates the fact that the particular choice of index for the free subscript in an equation is not important provided that the same free subscript appears in every symbol grouping.[1]

As a second example, suppose that $f(x_1, x_2, \ldots, x_n)$ is a function of $x_1, x_2, \ldots, x_n$, Then, if we write the equation

$$\frac{\partial f}{\partial x_k} = 3x_k, \tag{1.28}$$

the index k in it is a free subscript and so takes all values in the range $1, 2, \ldots, n$. Thus (1.28) is a compact way of writing the n equations

$$\frac{\partial f}{\partial x_1} = 3x_1, \qquad \frac{\partial f}{\partial x_2} = 3x_2, \qquad \ldots, \qquad \frac{\partial f}{\partial x_n} = 3x_n. \tag{1.29}$$

As a third example, the equation

$$A_{pq} = x_p x_q \tag{1.30}$$

has two free subscripts p and q, and each, independently, takes all values in the range $1, 2, \ldots, n$. Therefore (1.30) corresponds to the nine equations

$$\left.\begin{array}{llll} A_{11} = x_1x_1, & A_{12} = x_1x_2, & \ldots & A_{1n} = x_1x_n, \\ A_{21} = x_2x_1, & A_{22} = x_2x_2, & \ldots & A_{2n} = x_2x_n, \\ \ldots & \ldots & \ldots & \ldots \quad = \ldots \\ A_{n1} = x_nx_1, & A_{n2} = x_nx_2, & \ldots & A_{nn} = x_nx_n. \end{array}\right\} \tag{1.31}$$

In general, if an equation involves N free indices, then it represents 3^N scalar equations.

In order to be consistent it is important that *the same free subscript(s) must appear once, and only once, in every group of symbols in an equation.* For example, in equation (1.26), since the index i appears once in the symbol group $A_{i1}x_1$, it must necessarily appear once in each of the remaining symbol groups $A_{i2}x_2$, $A_{i3}x_3, \ldots$ $A_{in}x_n$ and b_i of that equation. Similarly since the free subscripts p and q appear in the symbol group on the left-hand

[1] By a "symbol group" we mean a set of terms contained between $+, -$ and $=$ signs.

side of equation (1.30), it must also appear in the symbol group on the right-hand side. An equation of the form $A_{pq} = x_i x_j$ would violate this consistency requirement as would $A_{i1}x_i + A_{j2}x_2 = 0$.

Note finally that had we adopted the range convention in Section 1.1, we would have omitted the various "i=1,2,. . . ,n" statements there and written, for example, equation (1.4) for the equality of two matrices as simply $A_{ij} = B_{ij}$; equation (1.5) for the sum of two matrices as simply $C_{ij} = A_{ij} + B_{ij}$; equation (1.7) for the scalar multiple of a matrix as $B_{ij} = \alpha A_{ij}$; equation (1.8) for the transpose of a matrix as simply $B_{ij} = A_{ji}$; equation (1.11) defining a symmetric matrix as simply $A_{ij} = A_{ji}$; and equation (1.12) defining a skew-symmetric matrix as simply $A_{ij} = -A_{ji}$.

1.3 Summation convention

Next, observe that (1.26) can be written as

$$\sum_{j=1}^{n} A_{ij}x_j = b_i. \tag{1.32}$$

We can simplify the notation even further *by agreeing to drop the summation sign and instead imposing the rule that summation is implied over a subscript that appears twice in a symbol grouping.* With this understanding in force, we would write (1.32) as

$$A_{ij}x_j = b_i \tag{1.33}$$

with summation on the subscript j being implied. A subscript that appears twice in a symbol grouping is called a *repeated* or *dummy* subscript; the subscript j in (1.33) is a dummy subscript.

Note that

$$A_{ik}x_k = b_i \tag{1.34}$$

is identical to (1.33); this is because k is a dummy subscript in (1.34) and therefore summation on k in implied in (1.34). Thus the particular choice of index for the dummy subscript is not important.

In order to avoid ambiguity, no subscript is allowed to appear more than twice in any symbol grouping. Thus we shall never write, for example, $A_{ii}x_i = b_i$ since, if we did, the index i would appear 3 times in the first symbol group.

Summary of Rules:

1. Lower-case latin subscripts take on values in the range $(1, 2, \ldots, n)$.
2. A given index may appear either once or twice in a symbol grouping. If it appears once, it is called a free index and it takes on each value in its range. If it appears twice, it is called a dummy index and summation is implied over it.
3. The same index may not appear more than twice in the same symbol grouping.
4. All symbol groupings in an equation must have the same free subscripts.

Free and dummy indices may be changed without altering the meaning of an expression *provided that* one does not violate the preceding rules. Thus, for example, we can change the free subscript p in every term of the equation

$$A_{pq} x_q \;=\; b_p \tag{1.35}$$

to any other index, say k, and equivalently write

$$A_{kq} x_q \;=\; b_k. \tag{1.36}$$

We can also change the repeated subscript q to some other index, say s, and write

$$A_{ks} x_s \;=\; b_k. \tag{1.37}$$

The three preceding equations are identical.

It is important to emphasize that each of the equations in, for example (1.24), involves scalar quantities, and therefore, the order in which the terms appear within a symbol group is irrelevant. Thus, for example, $(1.24)_1$ is equivalent to $x_1 A_{11} + x_2 A_{12} + \ldots + x_n A_{1n} = b_1$. Likewise we can write (1.33) equivalently as $x_j A_{ij} = b_i$. Note that both $A_{ij} x_j = b_i$ and $x_j A_{ij} = b_i$ represent the matrix equation $[A]\{x\} = \{b\}$; the second equation does *not* correspond to $\{x\}[A] = \{b\}$. In an indicial equation it is the location of the subscripts that is crucial; in particular, it is the location where the repeated subscript appears that tells us whether $\{x\}$ multiplies $[A]$ or $[A]$ multiplies $\{x\}$.

Note finally that had we adopted the range and summation conventions in Section 1.1, we would have written equation (1.6) for the product of two matrices as $C_{ij} = A_{ik} B_{kj}$; equation (1.10) for the product of a matrix by its transpose as $\{x\}^T\{x\} = x_i x_i$; equation (1.13) for the quadratic form as $\{x\}^T[A]\{x\} = A_{ij} x_i x_j$; and equation (1.15) for the trace as trace $[A] = A_{ii}$.

1.4 Kronecker delta

The *Kronecker Delta*, δ_{ij}, is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j. \end{cases} \tag{1.38}$$

Note that it represents the elements of the identity matrix. If $[Q]$ is an orthogonal matrix, then we know that $[Q][Q]^T = [Q]^T[Q] = [I]$. This implies, in indicial notation, that

$$Q_{ik}Q_{jk} = Q_{ki}Q_{kj} = \delta_{ij} \ . \tag{1.39}$$

The following useful property of the Kronecker delta is sometimes called the *substitution rule*. Consider, for example, any column matrix $\{u\}$ and suppose that one wishes to simplify the expression $u_i\delta_{ij}$. Recall that $u_i\delta_{ij} = u_1\delta_{1j} + u_2\delta_{2j} + \ldots + u_n\delta_{nj}$. Since δ_{ij} is zero unless $i = j$, it follows that all terms on the right-hand side vanish trivially except for the one term for which $i = j$. Thus the term that survives on the right-hand side is u_j and so

$$u_i\delta_{ij} = u_j. \tag{1.40}$$

Thus we have used the facts that *(i)* since δ_{ij} is zero unless $i = j$, the expression being simplified has a non-zero value only if $i = j$; *(ii)* and when $i = j$, δ_{ij} is unity. Thus replacing the Kronecker delta by unity, and changing the repeated subscript $i \rightarrow j$, gives $u_i\delta_{ij} = u_j$. Similarly, suppose that $[A]$ is a square matrix and one wishes to simplify $A_{jk}\delta_{\ell j}$. Then by the same reasoning, we replace the Kronecker delta by unity and change the repeated subscript $j \rightarrow \ell$ to obtain[2]

$$A_{jk}\delta_{\ell j} = A_{\ell k}. \tag{1.41}$$

More generally, if δ_{ip} multiplies a quantity $\mathbb{C}_{ij\ell k}$ representing n^4 numbers, one replaces the Kronecker delta by unity and changes the repeated subscript $i \rightarrow p$ to obtain

$$\mathbb{C}_{ij\ell k} \ \delta_{ip} = \mathbb{C}_{pj\ell k}. \tag{1.42}$$

The substitution rule applies even more generally: for any quantity or expression $T_{ipq\ldots z}$, one simply replaces the Kronecker delta by unity and changes the repeated subscript $i \rightarrow j$ to obtain

$$T_{ipq\ldots z} \ \delta_{ij} = T_{jpq\ldots z}. \tag{1.43}$$

[2] Observe that these results are immediately apparent by using matrix algebra. In the first example, note that $\delta_{ji}u_i$ (which is equal to the quantity $\delta_{ij}u_i$ that is given) is simply the jth element of the column matrix $[I]\{u\}$. Since $[I]\{u\} = \{u\}$ the result follows at once. Similarly in the second example, $\delta_{\ell j}A_{jk}$ is simply the ℓ, k-element of the matrix $[I][A]$. Since $[I][A] = [A]$, the result follows.

1.5 The alternator or permutation symbol

We now limit attention to subscripts that *range over* 1, 2, 3 *only.* The *alternator or permutation symbol* is defined by

$$\begin{aligned} e_{ijk} &= \begin{cases} 0 & \text{if two or more subscripts i, j, k, are equal,} \\ +1 & \text{if the subscripts i, j, k, are in cyclic order,} \\ -1 & \text{if the subscripts i, j, k, are in anticyclic order,} \end{cases} \\ &= \begin{cases} 0 & \text{if two or more subscripts i, j, k, are equal,} \\ +1 & \text{for } (i,j,k) = (1,2,3),(2,3,1),(3,1,2), \\ -1 & \text{for } (i,j,k) = (1,3,2),(2,1,3),(3,2,1). \end{cases} \end{aligned} \tag{1.44}$$

Observe from its definition that the sign of e_{ijk} changes whenever any two adjacent subscripts are switched:

$$e_{ijk} = -e_{jik} = e_{jki}. \tag{1.45}$$

One can show by direct calculation that the determinant of a 3 matrix $[A]$ can be written in either of two forms

$$\det[A] = e_{ijk}A_{1i}A_{2j}A_{3k} \qquad \text{or} \qquad \det[A] = e_{ijk}A_{i1}A_{j2}A_{k3}; \tag{1.46}$$

as well as in the form

$$\det[A] = \frac{1}{6}\, e_{ijk}e_{pqr}A_{ip}A_{jq}A_{kr}. \tag{1.47}$$

Another useful identity involving the determinant is

$$e_{pqr}\det[A] = e_{ijk}A_{ip}A_{jq}A_{kr}. \tag{1.48}$$

The following relation involving the alternator and the Kronecker delta will be useful in subsequent calculations

$$e_{ijk}e_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}. \tag{1.49}$$

It is left to the reader to develop proofs of these identities. They can, of course, be verified directly, by simply writing out all of the terms in (1.46) - (1.49).

1.6 Worked Examples.

Example(1.1): If $[A]$ and $[B]$ are $n \times n$ square matrices and $\{x\}, \{y\}, \{z\}$ are $n \times 1$ column matrices, express the matrix equation

$$\{y\} = [A]\{x\} + [B]\{z\}$$

as a set of scalar equations.

Solution: By the rules of matrix multiplication, the element y_i in the i^{th} row of $\{y\}$ is obtained by first pairwise multiplying the elements $A_{i1}, A_{i2}, \ldots, A_{in}$ of the i^{th} row of $[A]$ by the respective elements $x_1, x_2, \ldots, x_n$ of $\{x\}$ and summing; then doing the same for the elements of $[B]$ and $\{z\}$; and finally adding the two together. Thus

$$y_i = A_{ij} x_j + B_{ij} z_j \ ,$$

where summation over the dummy index j is implied, and this equation holds for each value of the free index $i = 1, 2, \ldots, n$. Note that one can alternatively – and *equivalently* – write the above equation in any of the following forms:

$$y_k = A_{kj} x_j + B_{kj} z_j, \qquad y_k = A_{kp} x_p + B_{kp} z_p, \qquad y_i = A_{ip} x_p + B_{iq} z_q.$$

Observe that all rules for indicial notation are satisfied by each of the three equations above.

Example(1.2): The $n \times n$ matrices $[C], [D]$ and $[E]$ are defined in terms of the two $n \times n$ matrices $[A]$ and $[B]$ by

$$[C] = [A][B], \qquad [D] = [B][A], \qquad [E] = [A][B]^T.$$

Express the elements of $[C], [D]$ and $[E]$ in terms of the elements of $[A]$ and $[B]$.

Solution: By the rules of matrix multiplication, the element C_{ij} in the i^{th} row and j^{th} column of $[C]$ is obtained by multiplying the elements of the i^{th} row of $[A]$, pairwise, by the respective elements of the j^{th} column of $[B]$ and summing. So, C_{ij} is obtained by multiplying the elements $A_{i1}, A_{i2}, \ldots A_{in}$ by, respectively, $B_{1j}, B_{2j}, \ldots B_{nj}$ and summing. Thus

$$C_{ij} = A_{ik} B_{kj};$$

note that i and j are both free indices here and so this represents n^2 scalar equations; moreover summation is carried out over the repeated index k. It follows likewise that the equation $[D] = [B][A]$ leads to

$$D_{ij} = B_{ik} A_{kj}; \qquad \text{or equivalently} \qquad D_{ij} = A_{kj} B_{ik},$$

where the second expression was obtained by simply changing the order in which the terms appear in the first expression (since, as noted previously, the order of terms within a symbol group is insignificant since these are scalar quantities.) In order to calculate E_{ij}, we first multiply $[A]$ by $[B]^T$ to obtain $E_{ij} = A_{ik} B^T_{kj}$. However, by definition of transposition, the i, j-element of a matrix $[B]^T$ equals the j, i-element of the matrix $[B]$: $B^T_{ij} = B_{ji}$ and so we can write

$$E_{ij} = A_{ik} B_{jk}.$$

All four expressions here involve the ik, kj or jk elements of $[A]$ and $[B]$. The precise locations of the subscripts vary and the meaning of the terms depend crucially on these locations. It is worth repeating that the location of the repeated subscript k tells us what term multiplies what term.

Example(1.3): If $[S]$ is any symmetric matrix and $[W]$ is any skew-symmetric matrix, show that

$$S_{ij}W_{ij} = 0.$$

Solution: Note that both i and j are dummy subscripts here; therefore there are summations over each of them. Also, there is no free subscript so this is just a single scalar equation.

Whenever there is a dummy subscript, the choice of the particular index for that dummy subscript is arbitrary, and we can change it to another index, provided that we change both repeated subscripts to the new symbol (and as long as we do not have any subscript appearing more than twice). Thus, for example, since i is a dummy subscript in $S_{ij}W_{ij}$, we can change $i \to p$ and get $S_{ij}W_{ij} = S_{pj}W_{pj}$. Note that we can change i to *any* other index except j; if we did change it to j, then there would be four j's and that violates one of our rules.

By changing the dummy indices $i \to p$ and $j \to q$, we get $S_{ij}W_{ij} = S_{pq}W_{pq}$. We can now change dummy indices again, from $p \to j$ and $q \to i$ which gives $S_{pq}W_{pq} = S_{ji}W_{ji}$. On combining, these we get

$$S_{ij}W_{ij} = S_{ji}W_{ji}.$$

Effectively, we have changed both i and j simultaneously from $i \to j$ and $j \to i$.

Next, since $[S]$ is symmetric $S_{ji} = S_{ij}$; and since $[W]$ is skew-symmetric, $W_{ji} = -W_{ij}$. Therefore $S_{ji}W_{ji} = -S_{ij}W_{ij}$. Using this in the right-hand side of the preceding equation gives

$$S_{ij}W_{ij} = -S_{ij}W_{ij}$$

from which it follows that $S_{ij}W_{ij} = 0$.

Remark: As a special case, take $S_{ij} = u_i u_j$ where $\{u\}$ is an arbitrary column matrix; note that this $[S]$ is symmetric. It follows that for *any* skew-symmetric $[W]$,

$$W_{ij}u_i u_j = 0 \quad \text{for all } u_i.$$

Example(1.4): Show that any matrix $[A]$ can be additively decomposed into the sum of a symmetric matrix and a skew-symmetric matrix.

Solution: Define matrices $[S]$ and $[W]$ in terms of the given the matrix $[A]$ as follows:

$$S_{ij} = \frac{1}{2}(A_{ij} + A_{ji}), \qquad W_{ij} = \frac{1}{2}(A_{ij} - A_{ji}).$$

It may be readily verified from these definitions that $S_{ij} = S_{ji}$ and that $W_{ij} = -W_{ij}$. Thus, the matrix $[S]$ is symmetric and $[W]$ is skew-symmetric. By adding the two equations in above one obtains

$$S_{ij} + W_{ij} = A_{ij},$$

or in matrix form, $[A] = [S] + [W]$.

Example (1.5): Show that the quadratic form $T_{ij}u_iu_j$ is unchanged if T_{ij} is replaced by its symmetric part. i.e. show that for any matrix $[T]$,

$$T_{ij}u_iu_j = S_{ij}u_iu_j \quad \text{for all } u_i \qquad \text{where } S_{ij} = \frac{1}{2}(T_{ij} + T_{ji}). \tag{i}$$

Solution: The result follows from the following calculation:

$$\begin{aligned} T_{ij}\, u_iu_j = \left(\frac{1}{2}T_{ij} + \frac{1}{2}T_{ij} + \frac{1}{2}T_{ji} - \frac{1}{2}T_{ji}\right) u_iu_j \quad &= \quad \frac{1}{2}(T_{ij} + T_{ji})\, u_iu_j + \frac{1}{2}(T_{ij} - T_{ji})\, u_iu_j \\ &= \quad S_{ij}\, u_iu_j, \end{aligned}$$

where in the last step we have used the facts that $A_{ij} = T_{ij} - T_{ji}$ is skew-symmetric, that $B_{ij} = u_iu_j$ is symmetric, and that for any symmetric matrix $[A]$ and any skew-symmetric matrix $[B]$, one has $A_{ij}B_{ij} = 0$.

Example (1.6): Suppose that $\mathbb{D}_{1111}, \mathbb{D}_{1112}, \ldots \mathbb{D}_{111n}, \ldots \mathbb{D}_{1121}, \mathbb{D}_{1122}, \ldots \mathbb{D}_{112n}, \ldots \mathbb{D}_{nnnn}$ are n^4 constants; and let $\mathbb{D}_{ijk\ell}$ denote a generic element of this set where each of the subscripts i, j, k, ℓ take all values in the range $1, 2, \ldots n$. Let $[E]$ be an arbitrary symmetric matrix and define the elements of a matrix $[A]$ by $A_{ij} = \mathbb{D}_{ijk\ell}E_{k\ell}$. Show that $[A]$ is unchanged if $\mathbb{D}_{ijk\ell}$ is replaced by its "symmetric part" $\mathbb{C}_{ijk\ell}$ where

$$\mathbb{C}_{ijk\ell} = \frac{1}{2}(\mathbb{D}_{ijk\ell} + \mathbb{D}_{ij\ell k}). \tag{i}$$

Solution: In a manner entirely analogous to the previous example,

$$\begin{aligned} A_{ij} = \mathbb{D}_{ijk\ell}E_{k\ell} \quad &= \quad \left(\frac{1}{2}\mathbb{D}_{ijk\ell} + \frac{1}{2}\mathbb{D}_{ijk\ell} + \frac{1}{2}\mathbb{D}_{ij\ell k} - \frac{1}{2}\mathbb{D}_{ij\ell k}\right) E_{k\ell} \\ &= \quad \frac{1}{2}(\mathbb{D}_{ijk\ell} + \mathbb{D}_{ij\ell k})\, E_{k\ell} + \frac{1}{2}(\mathbb{D}_{ijk\ell} - \mathbb{D}_{ij\ell k})\, E_{k\ell} \\ &= \quad \mathbb{C}_{ijk\ell}\, E_{k\ell}, \end{aligned}$$

where in the last step we have used the fact that $(\mathbb{D}_{ijk\ell} - \mathbb{D}_{ij\ell k})E_{k\ell} = 0$ since $\mathbb{D}_{ijk\ell} - \mathbb{D}_{ij\ell k}$ is skew symmetric in the subscripts k, ℓ while $E_{k\ell}$ is symmetric in the subscripts k, ℓ.

Example (1.7): Evaluate the expression $\delta_{ij}\delta_{ik}\delta_{jk}$.

Solution: By using the substitution rule, first on the repeated index i and then on the repeated index j, we have $\delta_{ij}\, \delta_{ik}\, \delta_{jk} = \delta_{jk}\, \delta_{jk} = \delta_{kk} = \delta_{11} + \delta_{22} + \ldots + \delta_{nn} = n$.

Example(1.8): Given an orthogonal matrix $[Q]$, use indicial notation to solve the matrix equation $[Q]\{x\} = \{a\}$ for $\{x\}$.

Solution: In indicial form, the equation $[Q]\{x\} = \{a\}$ reads

$$Q_{ij}x_j = a_i.$$

Multiplying both sides by Q_{ik} gives

$$Q_{ik}Q_{ij}x_j = Q_{ik}a_i.$$

Since $[Q]$ is orthogonal, we know from (1.39) that $Q_{rp}Q_{rq} = \delta_{pq}$. Thus the preceding equation simplifies to

$$\delta_{jk}x_j = Q_{ik}a_i,$$

which, by the substitution rule, reduces further to

$$x_k = Q_{ik}a_i\ .$$

In matrix notation this reads $\{x\} = [Q]^T\{a\}$ which we could, of course, have written down immediately from the fact that $\{x\} = [Q]^{-1}\{a\}$, and for an orthogonal matrix, $[Q]^{-1} = [Q]^T$.

Example(1.9): Consider the function $f(x_1, x_2, \ldots, x_n) = A_{ij}x_ix_j$ where the A_{ij}'s are constants. Calculate the partial derivatives $\partial f/\partial x_i$.

Solution: We begin by making two general observations. First, note that because of the summation on the indices i and j, it is incorrect to conclude that $\partial f/\partial x_i = A_{ij}x_j$ by viewing this in the same way as differentiating the function $A_{12}x_1x_2$ with respect to x_1. Second, observe that if we differentiatiate f with respect to x_i and write $\partial f/\partial x_i = \partial(A_{ij}x_ix_j)/\partial x_i$, we would violate our rules because the right-hand side has the subscript i appearing three times in one symbol grouping. In order to get around this difficulty we make use of the fact that the specific choice of the index in a dummy subscript is not significant and so we can write $f = A_{pq}x_px_q$.

Differentiating f and using the fact that $[A]$ is constant gives

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}(A_{pq}x_px_q) = A_{pq}\frac{\partial}{\partial x_i}(x_px_q) = A_{pq}\left[\frac{\partial x_p}{\partial x_i}\ x_q + x_p\frac{\partial x_q}{\partial x_i}\right]\ .$$

Since the x_i's are independent variables, it follows that

$$\frac{\partial x_i}{\partial x_j} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases} \qquad i.e.\ \frac{\partial x_i}{\partial x_j}\ =\ \delta_{ij}\ .$$

Using this above gives

$$\frac{\partial f}{\partial x_i} = A_{pq}\left[\delta_{pi}x_q + x_p\delta_{qi}\right] = A_{pq}\delta_{pi}x_q + A_{pq}x_p\delta_{qi}$$

which, by the substitution rule, simplifies to

$$\frac{\partial f}{\partial x_i} = A_{iq}x_q + A_{pi}x_p = A_{ij}x_j + A_{ji}x_j = (A_{ij} + A_{ji})x_j\ .$$

Example (1.10): Suppose that $\{x\}^T[A]\{x\} = 0$ for *all* column matrices $\{x\}$ where the square matrix $[A]$ is independent of $\{x\}$. What does this imply about $[A]$?

Solution: We know from a previous example that that if $[A]$ is a skew-symmetric and $[S]$ is symmetric then $A_{ij}S_{ij} = 0$, and as a special case of this that $A_{ij}x_ix_j = 0$ for all $\{x\}$. Thus a sufficient condition for the given equation to hold is that $[A]$ be skew-symmetric. Now we show that this is also a necessary condition.

We are given that $A_{ij}x_ix_j = 0$ for all x_i. Since this equation holds for *all* x_i, we may differentiate both sides with respect to x_k and proceed as follows:

$$0 = \frac{\partial}{\partial x_k}(A_{ij}x_ix_j) = A_{ij}\frac{\partial}{\partial x_k}(x_ix_j) = A_{ij}\frac{\partial x_i}{\partial x_k}\,x_j + A_{ij}x_i\,\frac{\partial x_j}{\partial x_k} = A_{ij}\delta_{ik}\,x_j + A_{ij}\,x_i\,\delta_{jk}\ , \tag{i}$$

where we have used the fact that $\partial x_i/\partial x_j = \delta_{ij}$ in the last step. On using the substitution rule, this simplifies to

$$A_{kj}\,x_j + A_{ik}\,x_i = (A_{kj} + A_{jk})\,x_j = 0. \tag{ii}$$

Since this also holds for all x_i, it may be differentiated again with respect to x_i to obtain

$$(A_{kj} + A_{jk})\,\frac{\partial x_j}{\partial x_i} = (A_{kj} + A_{jk})\,\delta_{ji} = A_{ki} + A_{ik} = 0. \tag{iii}$$

Thus $[A]$ must necessarily be a skew symmetric matrix,

Therefore it is necessary and sufficient that $[A]$ be skew-symmetric.

Example (1.11): Let $\mathbb{C}_{ijkl}$ be a set of n^4 constants. Define the function $\widehat{W}([E])$ for all matrices $[E]$ by $\widehat{W}([E]) = W(E_{11}, E_{12}, ...E_{nn}) \ = \ \frac{1}{2}\,\mathbb{C}_{ijkl}E_{ij}E_{kl}$. Calculate

$$\frac{\partial W}{\partial E_{ij}} \qquad \text{and} \qquad \frac{\partial^2 W}{\partial E_{ij}\partial E_{kl}}. \tag{i}$$

Solution: First, since the E_{ij}'s are independent variables, it follows that

$$\frac{\partial E_{pq}}{\partial E_{ij}} \quad = \quad \begin{cases} 1 & \text{if } p = i \ \text{ and } \ q = j, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\frac{\partial E_{pq}}{\partial E_{ij}} \ = \ \delta_{pi}\,\delta_{qj}. \tag{ii}$$

Keeping this in mind and differentiating $W(E_{11}, E_{12}, ...E_{33})$ with respect to E_{ij} gives

$$\begin{aligned} \frac{\partial W}{\partial E_{ij}} = \frac{\partial}{\partial E_{ij}}\left(\frac{1}{2}\,\mathbb{C}_{pqrs}E_{pq}E_{rs}\right) \ &= \ \frac{1}{2}\,\mathbb{C}_{pqrs}\left(\frac{\partial E_{pq}}{\partial E_{ij}}\,E_{rs} + E_{pq}\,\frac{\partial E_{rs}}{\partial E_{ij}}\right) \\ &= \ \frac{1}{2}\,\mathbb{C}_{pqrs}\ (\delta_{pi}\,\delta_{qj}\,E_{rs} + \delta_{ri}\,\delta_{sj}\,E_{pq}) \\ &= \ \frac{1}{2}\,\mathbb{C}_{ijrs}\,E_{rs} + \frac{1}{2}\,\mathbb{C}_{pqij}\,E_{pq} \\ &= \ \frac{1}{2}\,(\mathbb{C}_{ijpq} + \mathbb{C}_{pqij})\,E_{pq}. \end{aligned}$$

where we have made use of the substitution rule. (Note that in the first step we wrote $W = \frac{1}{2}\,\mathbb{C}_{pqrs}E_{pq}E_{rs}$ rather than $W = \frac{1}{2}\,\mathbb{C}_{ijkl}E_{ij}E_{kl}$ because we would violate our rules for indices had we written $\partial(\frac{1}{2}\,\mathbb{C}_{ijkl}E_{ij}E_{kl})/\partial E_{ij}$.) Differentiating this once more with respect to E_{kl} gives

$$\frac{\partial^2 W}{\partial E_{ij}\,\partial E_{kl}} = \frac{\partial}{\partial E_{k\ell}}\left(\frac{1}{2}\left(\mathbb{C}_{ijpq} + \mathbb{C}_{pqij}\right)E_{pq}\right) = \frac{1}{2}\left(\mathbb{C}_{ijpq} + \mathbb{C}_{pqij}\right)\delta_{pk}\delta_{ql} \qquad \text{(iii)}$$

$$= \frac{1}{2}\left(\mathbb{C}_{ijkl} + \mathbb{C}_{klij}\right) \qquad \text{(iv)}$$

Example (1.12): Evaluate the expression $e_{ijk}e_{kij}$.

Solution: By first using the skew symmetry property (1.45), then using the identity (1.49), and finally using the substitution rule, we have $e_{ijk}e_{kij} = -e_{ijk}e_{ikj} = -(\delta_{jk}\,\delta_{kj} - \delta_{jj}\,\delta_{kk}) = -(\delta_{jj} - \delta_{jj}\,\delta_{kk}) = -(3 - 3\times 3) = 6$.

Example(1.13): Show that

$$e_{ijk}S_{jk} = 0 \qquad \text{(i)}$$

if and only if the matrix $[S]$ is symmetric.

Solution: First, suppose that $[S]$ is symmetric. Pick and fix the free subscript i at any value $i = 1, 2, 3$. Then, we can think of e_{ijk} as the j, k element of a 3×3 matrix. Since $e_{ijk} = -e_{ikj}$ this is a skew-symmetric matrix. In a previous example we showed that $S_{ij}W_{ij} = 0$ for any symmetric matrix $[S]$ and any skew-symmetric matrix $[W]$. Consequently (i) must hold.

Conversely suppose that (i) holds for some matrix $[S]$. Multiplying (i) by e_{ipq} and using the identity (1.49) leads to

$$e_{ipq}e_{ijk}S_{jk} = (\delta_{pj}\delta_{qk} - \delta_{pk}\delta_{qj})S_{jk} = S_{pq} - S_{qp} = 0$$

where in the last step we have used the substitutin rule. Thus $S_{pq} = S_{qp}$ and so $[S]$ is symmetric.

Remark: Note as a special case of this result that

$$e_{ijk}v_jv_k = 0 \qquad \text{(ii)}$$

for any arbitrary column matrix $\{v\}$.

References

1. R.A. Frazer, W.J. Duncan and A.R. Collar, *Elementary Matrices*, Cambridge University Press, 1965.
2. R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, 1960.

Chapter 2

Vectors and Linear Transformations

Notation:

α	.....	scalar
$\mathbf{a}$	.....	vector
$\mathbf{A}$	.....	linear transformation

As mentioned in the Preface, Linear Algebra is a far richer subject than the very restricted glimpse provided here might suggest. The discussion in these notes is limited almost entirely to *(a)* real 3-dimensional Euclidean vector spaces, and *(b)* to linear transformations that carry vectors from one vector space into the same vector space. These notes are designed to *review* those aspects of linear algebra that will be encountered in our study of continuum mechanics; it is *not* meant to be a source for learning the subject of linear algebra for the first time.

The following notation will be consistently used: Greek letters will denote real numbers; lowercase boldface Latin letters will denote vectors; and uppercase boldface Latin letters will denote linear transformations. Thus, for example, $\alpha, \beta, \gamma ...$ will denote scalars (real numbers); $\mathbf{a}, \mathbf{b}, \mathbf{c}, ...$ will denote vectors; and $\mathbf{A}, \mathbf{B}, \mathbf{C}, ...$ will denote linear transformations. In particular, "$\mathbf{o}$" will denote the null vector while "$\mathbf{0}$" will denote the null linear transformation.

2.1 Vectors

A *vector space* V is a collection of elements, called vectors, together with two operations, addition and multiplication by a scalar. The operation of addition (has certain properties which we do not list here) and associates with each pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V, a vector denoted by $\mathbf{x}+\mathbf{y}$ that is also in V. In particular, it is assumed that there is a unique vector $\mathbf{o} \in \mathsf{V}$ called the null vector such that $\mathbf{x}+\mathbf{o}=\mathbf{x}$. The operation of scalar multiplication (has certain properties which we do not list here) and associates with each vector $\mathbf{x} \in \mathsf{V}$ and each real number α, another vector in V denoted by $\alpha\mathbf{x}$.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be k vectors in V. These vectors are said to be *linearly independent* if the only real numbers $\alpha_1, \alpha_2 \ldots, \alpha_k$ for which

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 \cdots + \alpha_k\mathbf{x}_k = \mathbf{o} \tag{2.1}$$

are the numbers $\alpha_1 = \alpha_2 = \ldots \alpha_k = 0$. If V contains n linearly independent vectors but does not contain $n+1$ linearly independent vectors, we say that the *dimension* of V is n. Unless stated otherwise, from *hereon we restrict attention to 3-dimensional vector spaces.*

If V is a vector space, any set of three linearly independent vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is said to be a *basis* for V. Given *any* vector $\mathbf{x} \in \mathsf{V}$ there exist a unique set of numbers ξ_1, ξ_2, ξ_3 such that

$$\mathbf{x} = \xi_1\mathbf{e}_1 + \xi_2\mathbf{e}_2 + \xi_3\mathbf{e}_3; \tag{2.2}$$

the numbers ξ_1, ξ_2, ξ_3 are called the *components* of $\mathbf{x}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

Let U be a subset of a vector space V; we say that U is a *subspace* (or linear manifold) of V if, for every $\mathbf{x}, \mathbf{y} \in \mathsf{U}$ and every real number α, the vectors $\mathbf{x}+\mathbf{y}$ and $\alpha\mathbf{x}$ are also in U. Thus a linear manifold U of V is itself a vector space under the same operations of addition and multiplication by a scalar as in V.

A *scalar-product* (or inner product or dot product) on V is a function which assigns to each pair of vectors $\mathbf{x}$, $\mathbf{y}$ in V a scalar, which we denote by $\mathbf{x} \cdot \mathbf{y}$. A scalar-product has certain properties which we do not list here except to note that it is required that

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x} \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathsf{V}. \tag{2.3}$$

A *Euclidean vector space* is a vector space together with an inner product on that space. From hereon we shall restrict attention to 3-dimensional Euclidean vector spaces and denote such a space by $\mathbb{E}_3$.

The *length* (or magnitude or norm) of a vector $\mathbf{x}$ is the scalar denoted by $|\mathbf{x}|$ and defined by

$$|\mathbf{x}| = (\mathbf{x}\cdot\mathbf{x})^{1/2}. \tag{2.4}$$

A vector has zero length if and only if it is the null vector. A *unit vector* is a vector of unit length. The *angle* θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\cos\ \theta = \frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}||\mathbf{y}|}, \qquad 0 \le \theta \le \pi. \tag{2.5}$$

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\mathbf{x}\cdot\mathbf{y} = 0$. It is obvious, nevertheless helpful, to note that if we are given two vectors $\mathbf{x}$ and $\mathbf{y}$ where $\mathbf{x}\cdot\mathbf{y} = 0$ and $\mathbf{y} \neq \mathbf{o}$, this does *not* necessarily imply that $\mathbf{x} = \mathbf{o}$; on the other hand if $\mathbf{x}\cdot\mathbf{y} = 0$ for *every* vector $\mathbf{y}$, then $\mathbf{x}$ must be the null vector.

An *orthonormal basis* is a triplet of mutually orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbb{E}_3$. For such a basis,

$$\mathbf{e}_i\cdot\mathbf{e}_j = \delta_{ij} \quad \text{for} \ \ i,j = 1,2,3, \tag{2.6}$$

where the Kronecker delta δ_{ij} is defined in the usual way by

$$\delta_{ij} = \begin{cases} 1 & \text{if} \ \ i = j, \\ 0 & \text{if} \ \ i \neq j. \end{cases} \tag{2.7}$$

A *vector-product* (or cross-product) on $\mathbb{E}_3$ is a function which assigns to each ordered pair of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{E}_3$, a vector, which we denote by $\mathbf{x}\times\ \mathbf{y}$. The vector-product must have certain properties (which we do not list here) except to note that it is required that

$$\mathbf{y}\times\mathbf{x} = -\mathbf{x}\times\mathbf{y} \qquad \text{for all} \quad \mathbf{x},\mathbf{y} \in \mathsf{V}. \tag{2.8}$$

One can show that

$$\mathbf{x}\times\mathbf{y} = |\mathbf{x}|\ |\mathbf{y}|\ \sin\theta\ \mathbf{n}, \tag{2.9}$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$ as defined by (2.5), and $\mathbf{n}$ is a unit vector in the direction $\mathbf{x}\times\mathbf{y}$ which therefore is normal to the plane defined by $\mathbf{x}$ and $\mathbf{y}$. Since $\mathbf{n}$ is parallel to $\mathbf{x}\times\mathbf{y}$, and since it has unit length, it follows that $\mathbf{n} = (\mathbf{x}\times\mathbf{y})/|(\mathbf{x}\times\mathbf{y})|$. The magnitude $|\mathbf{x}\times\ \mathbf{y}|$ of the cross-product can be interpreted geometrically as the area of the triangle formed by the vectors $\mathbf{x}$ and $\mathbf{y}$. A basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is said to be *right-handed* if

$$(\mathbf{e}_1\times\mathbf{e}_2)\cdot\mathbf{e}_3 > 0. \tag{2.10}$$

2.1.1 Euclidean point space

A *Euclidean point space* $\mathbb{P}$ whose elements are called points, is related to a Euclidean vector space $\mathbb{E}_3$ in the following manner. Every order pair of points (p, q) is uniquely associated with a vector in $\mathbb{E}_3$, say $\vec{pq}$, such that

(i) $\vec{pq} = \ - \vec{qp}$ for all $p, q \in \mathbb{P}$.

(ii) $\vec{pq} + \vec{qr} = \vec{pr}$ for all $p, q, r \in \mathbb{P}$.

(iii) given an arbitrary point $p \in \mathbb{P}$ and an arbitrary vector $\mathbf{x} \in \mathbb{E}_3$, there is a unique point $q \in \mathbb{P}$ such that $\mathbf{x} = \vec{pq}$. Here $\mathbf{x}$ is called the position of point q relative to the point p.

Pick and fix an arbitrary point $o \in \mathbb{P}$ (which we call the origin of $\mathbb{P}$) and an arbitrary basis for $\mathbb{E}_3$ of unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Corresponding to any point $p \in \mathbb{P}$ there is a unique vector $\vec{op} = \mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 \in \mathbb{E}_3$. The triplet (x_1, x_2, x_3) are called the coordinates of p in the *(coordinate) frame* $\mathcal{F} = \{o; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ comprised of the origin o and the basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. If $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is an orthonormal basis, the coordinate frame $\{o; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is called a *rectangular cartesian coordinate frame.*

2.2 Linear Transformations.

Consider a three-dimensional Euclidean vector space $\mathbb{E}_3$. Let $\mathbf{F}$ be a function (or transformation) which assigns to each vector $\mathbf{x} \in \mathbb{E}_3$, a second vector $\mathbf{y} \in \mathbb{E}_3$,

$$\mathbf{y} = \mathbf{F}(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{E}_3,\ \mathbf{y} \in \mathbb{E}_3; \tag{2.11}$$

$\mathbf{F}$ is said to be a *linear transformation* if it is such that

$$\mathbf{F}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{F}(\mathbf{x}) + \beta\mathbf{F}(\mathbf{y}) \tag{2.12}$$

for all scalars α, β and all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{E}_3$. When $\mathbf{F}$ is a linear transformation, we usually omit the parenthesis and write $\mathbf{Fx}$ instead of $\mathbf{F}(\mathbf{x})$. Note that $\mathbf{Fx}$ is a vector, and it is the image of $\mathbf{x}$ under the transformation $\mathbf{F}$.

A linear transformation is defined by the way it operates on vectors in $\mathbb{E}_3$. A geometric example of a linear transformation is the “projection operator” $\mathbf{\Pi}$ which projects vectors onto a given plane $\mathcal{P}$. Let $\mathcal{P}$ be the plane normal to the unit vector $\mathbf{n}$.; see Figure 2.1. For

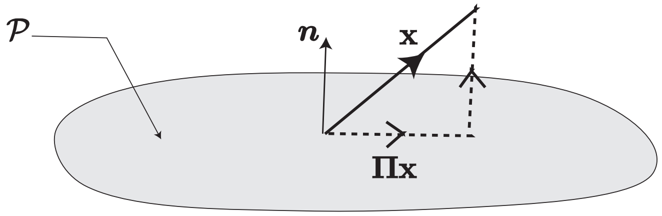


Figure 2.1: The projection $\mathbf{\Pi x}$ of a vector $\mathbf{x}$ onto the plane $\mathcal{P}$.

any vector $\mathbf{x} \in \mathbb{E}_3$, $\mathbf{\Pi x} \in \mathcal{P}$ is the vector obtained by projecting $\mathbf{x}$ onto $\mathcal{P}$. It can be verified geometrically that $\mathbf{P}$ is defined by

$$\mathbf{\Pi x} = \mathbf{x} - (\mathbf{x} \cdot \mathbf{n})\mathbf{n} \qquad \text{for all } \mathbf{x} \in \mathbb{E}_3. \tag{2.13}$$

Linear transformations tell us how vectors are mapped into other vectors. In particular, suppose that $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ are any three vectors in $\mathbb{E}_3$ and that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ are any three linearly independent vectors in $\mathbb{E}_3$. Then there is a unique linear transformation $\mathbf{F}$ that maps $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ into $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$: $\mathbf{y}_1 = \mathbf{F}\mathbf{x}_1, \mathbf{y}_2 = \mathbf{F}\mathbf{x}_2, \mathbf{y}_3 = \mathbf{F}\mathbf{x}_3$. This follows from the fact that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a basis for $\mathbb{E}_3$. Therefore any arbitrary vector $\mathbf{x}$ can be expressed uniquely in the form $\mathbf{x} = \xi_1\mathbf{x}_1 + \xi_2\mathbf{x}_2 + \xi_3\mathbf{x}_3$; consequently the image $\mathbf{F}\mathbf{x}$ of any vector $\mathbf{x}$ is given by $\mathbf{F}\mathbf{x} = \xi_1\mathbf{y}_1 + \xi_2\mathbf{y}_2 + \xi_3\mathbf{y}_3$ which is a rule for assigning a unique vector $\mathbf{F}\mathbf{x}$ to any given vector $\mathbf{x}$.

The *null linear transformation* $\mathbf{0}$ is the linear transformation that takes every vector $\mathbf{x}$ into the null vector $\mathbf{o}$. The *identity linear transformation* $\mathbf{I}$ takes every vector $\mathbf{x}$ into itself. Thus

$$\mathbf{0x} = \mathbf{o}, \qquad \mathbf{Ix} = \mathbf{x} \qquad \text{for all } \mathbf{x} \in \mathbb{E}_3. \tag{2.14}$$

Let $\mathbf{A}$ and $\mathbf{B}$ be linear transformations on $\mathbb{E}_3$ and let α be a scalar. The linear transformations $\mathbf{A}+\mathbf{B}$, $\mathbf{AB}$ and $\alpha\mathbf{A}$ are defined as those linear transformations which are such that

$$(\mathbf{A}+\mathbf{B})\mathbf{x} = \mathbf{Ax} + \mathbf{Bx} \quad \text{for all } \mathbf{x} \in \mathbb{E}_3, \tag{2.15}$$

$$(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx}) \qquad \text{for all } \mathbf{x} \in \mathbb{E}_3, \tag{2.16}$$

$$(\alpha\mathbf{A})\mathbf{x} = \alpha(\mathbf{Ax}) \qquad \text{for all } \mathbf{x} \in \mathbb{E}_3, \tag{2.17}$$

respectively; $\mathbf{A}+\mathbf{B}$ is called the *sum* of $\mathbf{A}$ and $\mathbf{B}, \mathbf{AB}$ the *product*, and $\alpha\mathbf{A}$ is the *scalar multiple* of $\mathbf{A}$ by α. In general,

$$\mathbf{AB} \neq \mathbf{BA}. \tag{2.18}$$

The *range* of a linear transformation $\mathbf{A}$ (i.e., the collection of all vectors $\mathbf{Ax}$ as $\mathbf{x}$ takes all values in $\mathbb{E}_3$) is a subspace of $\mathbb{E}_3$. The dimension of this particular subspace is known as the *rank* of $\mathbf{A}$. The set of all vectors $\mathbf{x}$ for which $\mathbf{Ax} = \mathbf{o}$ is also a subspace of $\mathbb{E}_3$; it is known as the *null space* of $\mathbf{A}$.

Given any linear transformation $\mathbf{A}$, one can show that there is a unique linear transformation usually denoted by $\mathbf{A}^T$ such that

$$\mathbf{Ax} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{A}^T\mathbf{y} \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{E}_3. \tag{2.19}$$

$\mathbf{A}^T$ is called the *transpose* of $\mathbf{A}$. One can show that

$$(\alpha\mathbf{A})^T = \alpha\mathbf{A}^T, \quad (\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T, \quad (\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T. \tag{2.20}$$

A linear transformation $\mathbf{A}$ is said to be *symmetric* if

$$\mathbf{A} = \mathbf{A}^T; \tag{2.21}$$

skew-symmetric if

$$\mathbf{A} = -\mathbf{A}^T. \tag{2.22}$$

Every linear transformation $\mathbf{A}$ can be represented as the sum of a symmetric linear transformation $\mathbf{S}$ and a skew-symmetric linear transformation $\mathbf{W}$ as follows:

$$\mathbf{A} = \mathbf{S} + \mathbf{W} \quad \text{where} \quad \mathbf{S} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T), \;\; \mathbf{W} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^T). \tag{2.23}$$

For every skew-symmetric linear transformation $\mathbf{W}$, it may be shown that

$$\mathbf{Wx} \cdot \mathbf{x} = 0 \qquad \text{for all } \mathbf{x} \in \mathbb{E}_3; \tag{2.24}$$

moreover, there exists a vector $\mathbf{w}$ (called the axial vector of $\mathbf{W}$) which has the property that

$$\mathbf{Wx} = \mathbf{w} \times \mathbf{x} \qquad \text{for all } \ \mathbf{x} \in \mathbb{E}_3. \tag{2.25}$$

Given a linear transformation $\mathbf{A}$, if the *only* vector $\mathbf{x}$ for which $\mathbf{Ax} = \mathbf{o}$ is the zero vector, then we say that $\mathbf{A}$ is *non-singular*. It follows from this that if $\mathbf{A}$ is non-singular then $\mathbf{Ax} \neq \mathbf{Ay}$ whenever $\mathbf{x} \neq \mathbf{y}$. Thus, a non-singular transformation $\mathbf{A}$ is a one-to-one transformation in the sense that, for any given $\mathbf{y} \in \mathbb{E}_3$, there is one and only one vector $\mathbf{x} \in \mathbb{E}_3$ for which $\mathbf{Ax} = \mathbf{y}$. Consequently, corresponding to any non-singular linear transformation

$\mathbf{A}$, there exists a second linear transformation, denoted by $\mathbf{A}^{-1}$ and called the *inverse* of $\mathbf{A}$, such that $\mathbf{A}\mathbf{x} = \mathbf{y}$ if and only if $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, or equivalently, such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}. \tag{2.26}$$

If $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ and $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ are two sets of linearly independent vectors in $\mathbb{E}_3$, then there is a unique non-singular linear transformation $\mathbf{F}$ that maps $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ into $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$: $\mathbf{y}_1 = \mathbf{F}\mathbf{x}_1, \mathbf{y}_2 = \mathbf{F}\mathbf{x}_2, \mathbf{y}_3 = \mathbf{F}\mathbf{x}_3$. The inverse of $\mathbf{F}$ maps $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ into $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$. If both bases $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ are right-handed (or both are left-handed) we say that the linear transformation $\mathbf{F}$ *preserves the orientation* of the vector space.

If two linear transformations $\mathbf{A}$ and $\mathbf{B}$ are both non-singular, then so is $\mathbf{AB}$; moreover,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{2.27}$$

If $\mathbf{A}$ is non-singular then so is $\mathbf{A}^T$; moreover,

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T, \tag{2.28}$$

and so there is no ambiguity in writing this linear transformation as $\mathbf{A}^{-T}$.

A linear transformation $\mathbf{Q}$ is said to be *orthogonal* if it preserves length, i.e., if

$$|\mathbf{Q}\mathbf{x}| = |\mathbf{x}| \text{ for all } \mathbf{x} \in \mathbb{E}_3. \tag{2.29}$$

If $\mathbf{Q}$ is orthogonal, it follows that it also preserves the inner product:

$$\mathbf{Q}\mathbf{x} \cdot \mathbf{Q}\mathbf{y} = \mathbf{x} \cdot \mathbf{y} \qquad \text{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{E}_3. \tag{2.30}$$

Thus an orthogonal linear transformation preserves both the length of a vector and the angle between two vectors. If $\mathbf{Q}$ is orthogonal, it is necessarily non-singular and

$$\mathbf{Q}^{-1} = \mathbf{Q}^T. \tag{2.31}$$

A linear transformation $\mathbf{A}$ is said to be *positive definite* if

$$\mathbf{A}\mathbf{x} \cdot \mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{E}_3, \ \mathbf{x} \neq \mathbf{o}; \tag{2.32}$$

positive-semi-definite if

$$\mathbf{A}\mathbf{x} \cdot \underline{\mathbf{x}} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{E}_3. \tag{2.33}$$

A positive definite linear transformation is necessarily non-singular. Moreover, $\mathbf{A}$ is positive definite if and only if its symmetric part $(1/2)(\mathbf{A}+\mathbf{A}^T)$ is positive definite.

Let $\mathbf{A}$ be a linear transformation. A subspace U is known as an *invariant subspace* of $\mathbf{A}$ if $\mathbf{Av} \in \mathsf{U}$ for all $\mathbf{v} \in \mathsf{U}$. Given a linear transformation $\mathbf{A}$, suppose that there exists an associated *one-dimensional* invariant subspace. Since U is one-dimensional, it follows that if $\mathbf{v} \in \mathsf{U}$ then any other vector in U can be expressed in the form $\lambda\mathbf{v}$ for some scalar λ. Since U is an invariant subspace we know in addition that $\mathbf{Av} \in \mathsf{U}$ whenever $\mathbf{v} \in \mathsf{U}$. Combining these two fact shows that $\mathbf{Av} = \lambda\mathbf{v}$ for all $\mathbf{v} \in \mathsf{U}$. A vector $\mathbf{v}$ and a scalar λ such that

$$\mathbf{Av} = \lambda\mathbf{v}, \tag{2.34}$$

are known, respectively, as an *eigenvector* and an *eigenvalue* of $\mathbf{A}$. Each eigenvector of $\mathbf{A}$ characterizes a one-dimensional invariant subspace of $\mathbf{A}$. Every linear transformation $\mathbf{A}$ (on a 3-dimensional vector space $\mathbb{E}_3$) has at least one eigenvalue.

It can be shown that a symmetric linear transformation $\mathbf{A}$ has three real eigenvalues λ_1, λ_2, and λ_3, and a corresponding set of three mutually orthogonal eigenvectors $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$. The particular basis of $\mathbb{E}_3$ comprised of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is said to be a *principal basis* of $\mathbf{A}$.

Every eigenvalue of a positive definite linear transformation must be positive, and no eigenvalue of a non-singular linear transformation can be zero. A symmetric linear transformation is positive definite if and only if all three of its eigenvalues are positive.

If $\mathbf{e}$ and λ are an eigenvector and eigenvalue of a linear transformation $\mathbf{A}$, then for any positive integer n, it is easily seen that $\mathbf{e}$ and λ^n are an eigenvector and an eigenvalue of $\mathbf{A}^n$ where $\mathbf{A}^n = \mathbf{AA}...(n \text{ times})..\mathbf{AA}$; this continues to be true for negative integers m provided $\mathbf{A}$ is non-singular and if by $\mathbf{A}^{-m}$ we mean $(\mathbf{A}^{-1})^m, \ m > 0$.

Finally, according to the *polar decomposition theorem*, given any non-singular linear transformation $\mathbf{F}$, there exists unique symmetric positive definite linear transformations $\mathbf{U}$ and $\mathbf{V}$ and a unique orthogonal linear transformation $\mathbf{R}$ such that

$$\mathbf{F} = \mathbf{RU} = \mathbf{VR}. \tag{2.35}$$

If λ and $\mathbf{r}$ are an eigenvalue and eigenvector of $\mathbf{U}$, then it can be readily shown that λ and $\mathbf{Rr}$ are an eigenvalue and eigenvector of $\mathbf{V}$.

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{E}_3$, their *tensor-product* is the linear transformation usually denoted by $\mathbf{a} \otimes \mathbf{b}$, which is such that

$$(\mathbf{a} \otimes \mathbf{b})\mathbf{x} = (\mathbf{x} \cdot \mathbf{b})\mathbf{a} \qquad \text{for all } \mathbf{x} \in \mathbb{E}_3. \tag{2.36}$$

Observe that for any $\mathbf{x} \in \mathbb{E}_3$, the vector $(\mathbf{a} \otimes \mathbf{b})\mathbf{x}$ is parallel to the vector $\mathbf{a}$. Thus the range of the linear transformation $\mathbf{a} \otimes \mathbf{b}$ is the one-dimensional subspace of $\mathbb{E}_3$ consisting of all vectors parallel to $\mathbf{a}$. The rank of the linear transformation $\mathbf{a} \otimes \mathbf{b}$ is thus unity.

For any vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and $\mathbf{d}$ it is easily shown that

$$(\mathbf{a} \otimes \mathbf{b})^T = \mathbf{b} \otimes \mathbf{a}, \qquad\qquad (\mathbf{a} \otimes \mathbf{b})(\mathbf{c} \otimes \mathbf{d}) \ = \ (\mathbf{b} \cdot \mathbf{c})(\mathbf{a} \otimes \mathbf{d}). \tag{2.37}$$

The product of a linear transformation $\mathbf{A}$ with the linear transformation $\mathbf{a} \otimes \mathbf{b}$ gives

$$\mathbf{A}(\mathbf{a} \otimes \mathbf{b}) = (\mathbf{A}\mathbf{a}) \otimes \mathbf{b}, \qquad (\mathbf{a} \otimes \mathbf{b})\mathbf{A} = \mathbf{a} \otimes (\mathbf{A}^T\mathbf{b}). \tag{2.38}$$

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be an orthonormal basis. Since this is a basis, any vector in $\mathbb{E}_3$, and therefore in particular each of the vectors $\mathbf{A}\mathbf{e}_1, \mathbf{A}\mathbf{e}_2, \mathbf{A}\mathbf{e}_3$, can be expressed as a unique linear combination of the basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. It follows that there exist unique real numbers A_{ij} such that

$$\mathbf{A}\mathbf{e}_j = \sum_{i=1}^{3} A_{ij}\mathbf{e}_i, \qquad j = 1, 2, 3, \tag{2.39}$$

where A_{ij} is the i^{th} component on the vector $\mathbf{A}\mathbf{e}_j$. They can equivalently be expressed as $A_{ij} = \mathbf{e}_i \cdot (\mathbf{A}\mathbf{e}_j)$. The linear transformation $\mathbf{A}$ can now be represented as

$$\mathbf{A} = \sum_{i=1}^{3} \sum_{j=1}^{3} A_{ij}(\mathbf{e}_i \otimes \mathbf{e}_j). \tag{2.40}$$

One refers to the A_{ij}'s as the *components of the linear transformation* $\mathbf{A}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Note that

$$\sum_{i=1}^{3} \mathbf{e}_i \otimes \mathbf{e}_i = \mathbf{I}, \qquad \sum_{i=1}^{3} (\mathbf{A}\mathbf{e}_i) \otimes \mathbf{e}_i = \mathbf{A}. \tag{2.41}$$

Let $\mathbf{S}$ be a symmetric linear transformation with eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and corresponding (mutually orthogonal unit) eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Since $\mathbf{S}\mathbf{e}_j = \lambda_j\mathbf{e}_j$ for each $j = 1, 2, 3$, it follows from (2.39) that the components of $\mathbf{S}$ in the principal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are $S_{11} = \lambda_1, S_{21} = S_{31} = 0; S_{12} = 0, S_{22} = \lambda_2, S_{32} = 0; S_{13} = S_{23} = 0, S_{33} = \lambda_3$. It follows from the general representation (2.40) that $\mathbf{S}$ admits the representation

$$\mathbf{S} = \sum_{i=1}^{3} \lambda_i \ (\mathbf{e}_i \otimes \mathbf{e}_i); \tag{2.42}$$

this is called the *spectral representation* of a symmetric linear transformation. It can be readily shown that, for any positive integer n,

$$\mathbf{S}^n = \sum_{i=1}^{3} \lambda_i^n \ (\mathbf{e}_i \otimes \mathbf{e}_i); \tag{2.43}$$

if $\mathbf{S}$ is symmetric and non-singular, then

$$\mathbf{S}^{-1} = \sum_{i=1}^{3} (1/\lambda_i) \ (\mathbf{e}_i \otimes \mathbf{e}_i). \tag{2.44}$$

If $\mathbf{S}$ is symmetric and positive definite, there is a unique symmetric positive definite linear transformation $\mathbf{T}$ such that $\mathbf{T}^2 = \mathbf{S}$. We call $\mathbf{T}$ the positive definite square root of $\mathbf{S}$ and denote it by $\mathbf{T} = \sqrt{\mathbf{S}}$. It is readily seen that

$$\sqrt{\mathbf{S}} = \sum_{i=i}^{3} \sqrt{\lambda_i} \ (\mathbf{e}_i \otimes \mathbf{e}_i). \tag{2.45}$$

2.3 Worked Examples.

Example 2.1: Given three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, show that

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}).$$

Solution: By the properties of the vector-product, the vector $(\mathbf{a} + \mathbf{b})$ is normal to the vector $(\mathbf{a} + \mathbf{b}) \times \mathbf{c}$. Thus

$$(\mathbf{a} + \mathbf{b}) \cdot [(\mathbf{a} + \mathbf{b}) \times \mathbf{c}] = 0.$$

On expanding this out one obtains

$$\mathbf{a} \cdot (\mathbf{a} \times \mathbf{c}) + \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \cdot (\mathbf{a} \times \mathbf{c}) + \mathbf{b} \cdot (\mathbf{b} \times \mathbf{c}) = 0.$$

Since $\mathbf{a}$ is normal to $(\mathbf{a} \times \mathbf{c})$, and $\mathbf{b}$ is normal to $(\mathbf{b} \times \mathbf{c})$, the first and last terms in this equation vanish. Finally, recall that $\mathbf{a} \times \mathbf{c} = -\mathbf{c} \times \mathbf{a}$. Thus the preceding equation simplifies to

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}).$$

This establishes the first part of the result. The second part is shown analogously.

Example 2.2: Show that a necessary and sufficient condition for three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in $\mathbb{E}_3$ – none of which is the null vector – to be linearly dependent is that $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0$.

<u>*Solution:*</u> To show necessity, suppose that the three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, are linearly dependent. It follows that

$$\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c} = \mathbf{o}$$

for some real numbers α, β, γ, at least one of which is non zero. Taking the vector-product of this equation with $\mathbf{c}$ and then taking the scalar-product of the result with $\mathbf{a}$ leads to

$$\beta\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0.$$

Analogous calculations with the other pairs of vectors, and keeping in mind that $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$, leads to

$$\alpha\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0, \qquad \beta\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0, \qquad \gamma\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0.$$

Since at least one of α, β, γ is non-zero it follows that necessarily $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{o}$.

To show sufficiency, let $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0$ and assume that $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent. We will show that this is a contradiction whence $\mathbf{a}, \mathbf{b}, \mathbf{c}$ must be linearly dependent. By the properties of the vector-product, the vector $\mathbf{b} \times \mathbf{c}$ is normal to the plane defined by the vectors $\mathbf{b}$ and $\mathbf{c}$. By assumption, $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0$, and this implies that $\mathbf{a}$ is normal to $\mathbf{b} \times \mathbf{c}$. Since we are in $\mathbb{E}_3$ this means that $\mathbf{a}$ must lie in the plane defined by $\mathbf{b}$ and $\mathbf{c}$. This means they cannot be linearly independent.

<u>*Example 2.3:*</u> Interpret the quantity $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ geometrically in terms of the volume of the tetrahedron defined by the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

<u>*Solution:*</u> Consider the tetrahedron formed by the three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ as depicted in Figure 2.2. Its volume $V_0 = \frac{1}{3}\, A_0\, h_0$ where A_0 is the area of its base and h_0 is its height.

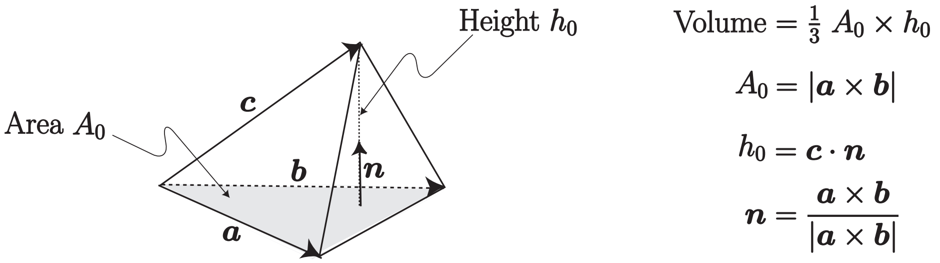

Height h_0
Area A_0
$\boldsymbol{c}$
$\boldsymbol{b}$
$\boldsymbol{n}$
$\boldsymbol{a}$
Volume $= \frac{1}{3}\, A_0 \times h_0$
$A_0 = |\boldsymbol{a} \times \boldsymbol{b}|$
$h_0 = \boldsymbol{c} \cdot \boldsymbol{n}$
$\boldsymbol{n} = \dfrac{\boldsymbol{a} \times \boldsymbol{b}}{|\boldsymbol{a} \times \boldsymbol{b}|}$

Figure 2.2: Volume of the tetrahedron defined by vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

Consider the triangle defined by the vectors $\mathbf{a}$ and $\mathbf{b}$ to be the base of the tetrahedron. Its area A_0 can be written as $1/2$ base × height $= 1/2|\mathbf{a}|(|\mathbf{b}||\sin\theta|)$ where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. However from the property (2.9) of the vector-product we have $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}||\sin\theta|$ and so $A_0 = |\mathbf{a} \times \mathbf{b}|/2$.

Next, $\mathbf{n} = (\mathbf{a} \times \mathbf{b})/|\mathbf{a} \times \mathbf{b}|$ is a unit vector that is normal to the base of the tetrahedron, and so the height of the tetrahedron is $h_0 = \mathbf{c} \cdot \mathbf{n}$; see Figure 2.2.

Therefore

$$V_0 = \frac{1}{3} A_0 h_0 = \frac{1}{3}\left(\frac{|\mathbf{a} \times \mathbf{b}|}{2}\right)(\mathbf{c} \cdot \mathbf{n}) = \frac{1}{6}\,(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}. \tag{i}$$

Observe that this provides a geometric explanation for why the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly dependent if and only if $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = 0$.

Example 2.4: Let $\phi(\mathbf{x})$ be a scalar-valued function defined on the vector space $\mathbb{E}_3$. If ϕ is linear, i.e. if $\phi(\alpha\mathbf{x}+\beta\mathbf{y}) = \alpha\phi(\mathbf{x})+\beta\phi(\mathbf{y})$ for all scalars α, β and all vectors $\mathbf{x}, \mathbf{y}$, show that $\phi(\mathbf{x}) = \mathbf{c}\cdot\mathbf{x}$ for some constant vector $\mathbf{c}$. Remark: This shows that the scalar-product is the most general scalar-valued linear function of a vector.

Solution: Let $\{\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_3\}$ be any orthonormal basis for $\mathbb{E}_3$. Then an arbitrary vector $\mathbf{x}$ can be written in terms of its components as $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$. Therefore

$$\phi(\mathbf{x}) = \phi(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3)$$

which because of the linearity of ϕ leads to

$$\phi(\mathbf{x}) = x_1\phi(\mathbf{e}_1) + x_2\phi(\mathbf{e}_2) + x_3\phi(\mathbf{e}_3).$$

On setting $c_i = \phi(\mathbf{e}_i), i = 1, 2, 3$, we find

$$\phi(\mathbf{x}) = x_1c_1 + x_2c_2 + x_3c_3 = \mathbf{c}\cdot\mathbf{x}$$

where $\mathbf{c} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3$.

Example 2.5: If two linear transformations $\mathbf{A}$ and $\mathbf{B}$ have the property that $\mathbf{Ax}\cdot\mathbf{y} = \mathbf{Bx}\cdot\mathbf{y}$ for all vectors $\mathbf{x}$ and $\mathbf{y}$, show that $\mathbf{A} = \mathbf{B}$.

Solution: Since $(\mathbf{Ax} - \mathbf{Bx})\cdot\mathbf{y} = 0$ for all vectors $\mathbf{y}$, we may choose $\mathbf{y} = \mathbf{Ax} - \mathbf{Bx}$ in this, leading to $|\mathbf{Ax} - \mathbf{Bx}|^2 = 0$. Since the only vector of zero length is the null vector, this implies that

$$\mathbf{Ax} = \mathbf{Bx} \qquad \text{for all vectors } \mathbf{x} \tag{i}$$

and so $\mathbf{A} = \mathbf{B}$.

Example 2.6: Let $\mathbf{n}$ be a unit vector, and let $\mathcal{P}$ be the plane through $\mathbf{o}$ normal to $\mathbf{n}$. Let $\mathbf{\Pi}$ and $\mathbf{R}$ be the transformations which, respectively, project and reflect a vector in the plane $\mathcal{P}$.

a. Show that $\mathbf{\Pi}$ and $\mathbf{R}$ are linear transformations; $\mathbf{\Pi}$ is called the "projection linear transformation" while $\mathbf{R}$ is known as the "reflection linear transformation".

b. Show that $\mathbf{R}(\mathbf{Rx}) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{E}_3$.

c. Verify that a reflection linear transformation $\mathbf{R}$ is non-singular while a projection linear transformation $\mathbf{\Pi}$ is singular. What is the inverse of $\mathbf{R}$?

d. Verify that a projection linear transformation $\mathbf{\Pi}$ is symmetric and that a reflection linear transformation $\mathbf{R}$ is orthogonal.

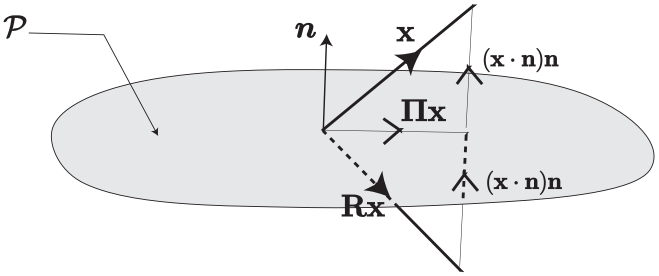

$\mathcal{P}$
$\boldsymbol{n}$
$\mathbf{x}$
$(\mathbf{x}\cdot\mathbf{n})\mathbf{n}$
$\mathbf{\Pi x}$
$(\mathbf{x}\cdot\mathbf{n})\mathbf{n}$
$\mathbf{Rx}$

Figure 2.3: The projection $\mathbf{\Pi x}$ and reflection $\mathbf{Rx}$ of a vector $\mathbf{x}$ on the plane $\mathcal{P}$.

e. Show that the projection linear transformation and reflection linear transformation can be represented as $\mathbf{\Pi} = \mathbf{I} - \mathbf{n}\otimes\mathbf{n}$ and $\mathbf{R} = \mathbf{I} - 2(\mathbf{n}\otimes\mathbf{n})$ respectively.

Solution:

a. Figure 2.3 shows a sketch of the plane $\mathcal{P}$, its unit normal vector $\mathbf{n}$, a generic vector $\mathbf{x}$, its projection $\mathbf{\Pi x}$ and its reflection $\mathbf{Rx}$. By geometry we see that

$$\mathbf{\Pi x} = \mathbf{x} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n}, \qquad \mathbf{Rx} = \mathbf{x} - 2(\mathbf{x}\cdot\mathbf{n})\mathbf{n}. \tag{i}$$

These define the images $\mathbf{\Pi x}$ and $\mathbf{Rx}$ of a generic vector $\mathbf{x}$ under the transformation $\mathbf{\Pi}$ and $\mathbf{R}$. One can readily verify that $\mathbf{\Pi}$ and $\mathbf{R}$ satisfy the requirement (2.12) of a linear transformation.

b. Applying the definition $(\mathrm{i})_2$ of $\mathbf{R}$ to the vector $\mathbf{Rx}$ gives

$$\mathbf{R}(\mathbf{Rx}) = (\mathbf{Rx}) - 2\Big((\mathbf{Rx})\cdot\mathbf{n}\Big)\mathbf{n}$$

Replacing $\mathbf{Rx}$ on the right-hand side of this equation by $(\mathrm{i})_2$, and expanding the resulting expression shows that the right-hand side simplifies to $\mathbf{x}$. Thus $\mathbf{R}(\mathbf{Rx}) = \mathbf{x}$.

c. Applying the definition $(\mathrm{i})_1$ of $\mathbf{\Pi}$ to the vector $\mathbf{n}$ gives

$$\mathbf{\Pi n} = \mathbf{n} - (\mathbf{n}\cdot\mathbf{n})\mathbf{n} = \mathbf{n} - \mathbf{n} = \mathbf{o}.$$

Therefore $\mathbf{\Pi n} = \mathbf{o}$ and (since $\mathbf{n} \neq \mathbf{o}$) we see that $\mathbf{o}$ is *not* the only vector that is mapped to the null vector by $\mathbf{\Pi}$. The transformation $\mathbf{\Pi}$ is therefore singular.

Next consider the transformation $\mathbf{R}$ and consider a vector $\mathbf{x}$ that is mapped by it to the null vector, i.e. $\mathbf{Rx} = \mathbf{o}$. Using $(\mathrm{i})_2$

$$\mathbf{x} = 2(\mathbf{x}\cdot\mathbf{n})\mathbf{n}.$$

Taking the scalar-product of this equation with the unit vector $\mathbf{n}$ yields $\mathbf{x}\cdot\mathbf{n} = 2(\mathbf{x}\cdot\mathbf{n})$ from which we conclude that $\mathbf{x}\cdot\mathbf{n} = 0$. Substituting this into the right-hand side of the preceding equation leads to $\mathbf{x} = \mathbf{o}$. Therefore $\mathbf{Rx} = \mathbf{o}$ if and only if $\mathbf{x} = \mathbf{o}$ and so $\mathbf{R}$ is non-singular.

To find the inverse of $\mathbf{R}$, recall from part (b) that $\mathbf{R}(\mathbf{Rx}) = \mathbf{x}$. Operating on both sides of this equation by $\mathbf{R}^{-1}$ gives $\mathbf{Rx} = \mathbf{R}^{-1}\mathbf{x}$. Since this holds for all vectors $\mathbf{x}$ it follows that $\mathbf{R}^{-1} = \mathbf{R}$.

d. To show that $\mathbf{\Pi}$ is symmetric we simply use its definition $(\mathrm{i})_1$ to calculate $\mathbf{\Pi x} \cdot \mathbf{y}$ and $\mathbf{x} \cdot \mathbf{\Pi y}$ for arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$. This yields

$$\mathbf{\Pi x} \cdot \mathbf{y} = \Big(\mathbf{x} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n}\Big) \cdot \mathbf{y} = \mathbf{x}\cdot\mathbf{y} - (\mathbf{x}\cdot\mathbf{n})(\mathbf{y}\cdot\mathbf{n})$$

and

$$\mathbf{x} \cdot \mathbf{\Pi y} = \mathbf{x} \cdot \Big(\mathbf{y} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n}\Big) = \mathbf{x}\cdot\mathbf{y} - (\mathbf{x}\cdot\mathbf{n})(\mathbf{y}\cdot\mathbf{n}).$$

Thus $\mathbf{\Pi x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{\Pi y}$ and so $\mathbf{\Pi}$ is symmetric.

To show that $\mathbf{R}$ is orthogonal we must show that $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ or $\mathbf{R}^T = \mathbf{R}^{-1}$. We begin by calculating $\mathbf{R}^T$. Recall from the definition (2.19) that the transpose satisfies the requirement $\mathbf{x}\cdot\mathbf{R}^T\mathbf{y} = \mathbf{R}\mathbf{x}\cdot\mathbf{y}$. Using the definition $(\mathrm{i})_2$ of $\mathbf{R}$ on the right-hand side of this equation yields

$$\mathbf{x}\cdot\mathbf{R}^T\mathbf{y} = \mathbf{x}\cdot\mathbf{y} - 2(\mathbf{x}\cdot\mathbf{n})(\mathbf{y}\cdot\mathbf{n}).$$

We can rearrange the right-hand side of this equation so it reads

$$\mathbf{x}\cdot\mathbf{R}^T\mathbf{y} = \mathbf{x}\cdot\Big(\mathbf{y} - 2(\mathbf{y}\cdot\mathbf{n})\mathbf{n}\Big).$$

Since this holds for all $\mathbf{x}$ it follows that $\mathbf{R}^T\mathbf{y} = \mathbf{y} - 2(\mathbf{y}\cdot\mathbf{n})\mathbf{n}$. Comparing this with $(\mathrm{i})_2$ shows that $\mathbf{R}^T = \mathbf{R}$. In part (c) we showed that $\mathbf{R}^{-1} = \mathbf{R}$ and so it now follows that $\mathbf{R}^T = \mathbf{R}^{-1}$. Thus $\mathbf{R}$ is orthogonal.

e. Applying the operation $(\mathbf{I} - \mathbf{n}\otimes\mathbf{n})$ on an arbitrary vector $\mathbf{x}$ gives

$$\Big(\mathbf{I} - \mathbf{n}\otimes\mathbf{n}\Big)\mathbf{x} = \mathbf{x} - (\mathbf{n}\otimes\mathbf{n})\mathbf{x} = \mathbf{x} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n} = \mathbf{\Pi x}$$

and so $\mathbf{\Pi} = \mathbf{I} - \mathbf{n}\otimes\mathbf{n}$.

Similarly

$$\Big(\mathbf{I} - 2\mathbf{n}\otimes\mathbf{n}\Big)\mathbf{x} = \mathbf{x} - 2(\mathbf{x}\cdot\mathbf{n})\mathbf{n} = \mathbf{R}\mathbf{x}$$

and so $\mathbf{R} = \mathbf{I} - 2\mathbf{n}\otimes\mathbf{n}$.

Example 2.7: If $\mathbf{W}$ is a skew symmetric linear transformation show that

$$\mathbf{W}\mathbf{x}\cdot\mathbf{x} = 0 \qquad \text{for all } \mathbf{x}\ . \tag{i}$$

Solution: By the definition (2.19) of the transpose, we have $\mathbf{W}\mathbf{x}\cdot\mathbf{x} = \mathbf{x}\cdot\mathbf{W}^T\mathbf{x}$; and since $\mathbf{W} = -\mathbf{W}^T$ for a skew symmetric linear transformation, this can be written as $\mathbf{W}\mathbf{x}\cdot\mathbf{x} = -\mathbf{x}\cdot\mathbf{W}\mathbf{x}$. Finally the property (2.3) of the scalar-product allows this to be written as $\mathbf{W}\mathbf{x}\cdot\mathbf{x} = -\mathbf{W}\mathbf{x}\cdot\mathbf{x}$ from which the desired result follows.

Example 2.8: Show that $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$.

Solution: First, by the definition (2.19) of the transpose,

$$(\mathbf{A}\mathbf{B})\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot(\mathbf{A}\mathbf{B})^T\mathbf{y}\ . \tag{i}$$

Second, note that $(\mathbf{AB})\mathbf{x}\cdot\mathbf{y} = \mathbf{A}(\mathbf{Bx})\cdot\mathbf{y}$. By the definition of the transpose of $\mathbf{A}$ we have $\mathbf{A}(\mathbf{Bx})\cdot\mathbf{y} = \mathbf{Bx}\cdot\mathbf{A}^T\mathbf{y}$; and by the definition of the transpose of $\mathbf{B}$ we have $\mathbf{Bx}\cdot\mathbf{A}^T\mathbf{y} = \mathbf{x}\cdot\mathbf{B}^T\mathbf{A}^T\mathbf{y}$. Therefore combining these three equations shows that

$$(\mathbf{AB})\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot\mathbf{B}^T\mathbf{A}^T\mathbf{y} \qquad \text{(ii)}$$

On equating these two expressions for $(\mathbf{AB})\mathbf{x}\cdot\mathbf{y}$ shows that $\mathbf{x}\cdot(\mathbf{AB})^T\mathbf{y} = \mathbf{x}\cdot\mathbf{B}^T\mathbf{A}^T\mathbf{y}$ for all vectors $\mathbf{x},\mathbf{y}$ which establishes the desired result.

Example 2.9: If $\mathbf{o}$ is the null vector, then show that $\mathbf{Ao} = \mathbf{o}$ for any linear transformation $\mathbf{A}$.

Solution: The null vector $\mathbf{o}$ has the property that when it is added to any vector, the vector remains unchanged. Therefore $\mathbf{x}+\mathbf{o} = \mathbf{x}$, and similarly $\mathbf{Ax}+\mathbf{o} = \mathbf{Ax}$. However operating on the first of these equations by $\mathbf{A}$ shows that $\mathbf{Ax}+\mathbf{Ao} = \mathbf{Ax}$, which when combined with the second equation yields the desired result.

Example 2.10: If $\mathbf{A}$ and $\mathbf{B}$ are non-singular linear transformations show that $\mathbf{AB}$ is also non-singular and that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Solution: Let $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. We will show that $(\mathbf{AB})\mathbf{C} = \mathbf{C}(\mathbf{AB}) = \mathbf{I}$ and therefore that $\mathbf{C}$ is the inverse of $\mathbf{AB}$. (Since the inverse would thus have been shown to exist, necessarily $\mathbf{AB}$ must be non-singular.)

Observe first that

$$(\mathbf{AB})\,\mathbf{C} = (\mathbf{AB})\,\mathbf{B}^{-1}\mathbf{A}^{-1} = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{I}\ ,$$

and similarly that

$$\mathbf{C}\,(\mathbf{AB}) = \mathbf{B}^{-1}\mathbf{A}^{-1}\,(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} == \mathbf{B}^{-1}\mathbf{IB} = \mathbf{I}\ .$$

Therefore $(\mathbf{AB})\mathbf{C} = \mathbf{C}(\mathbf{AB}) = \mathbf{I}$ and so $\mathbf{C}$ is the inverse of $\mathbf{AB}$.

Example 2.11: If $\mathbf{A}$ is non-singular, show that $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$.

Solution: Since $(\mathbf{A}^T)^{-1}$ is the inverse of $\mathbf{A}^T$ we have $(\mathbf{A}^T)^{-1}\mathbf{A}^T = \mathbf{I}$. Post-operating on both sides of this equation by $(\mathbf{A}^{-1})^T$ gives

$$(\mathbf{A}^T)^{-1}\mathbf{A}^T(\mathbf{A}^{-1})^T = (\mathbf{A}^{-1})^T.$$

Recall that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ for any two linear transformations $\mathbf{A}$ and $\mathbf{B}$. Thus the preceding equation simplifies to

$$(\mathbf{A}^T)^{-1}(\mathbf{A}^{-1}\mathbf{A})^T = (\mathbf{A}^{-1})^T$$

Since $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ the desired result follows.

Example 2.12: Show that an orthogonal linear transformation $\mathbf{Q}$ preserves inner products, i.e. show that $\mathbf{Qx}\cdot\mathbf{Qy} = \mathbf{x}\cdot\mathbf{y}$ for all vectors $\mathbf{x},\mathbf{y}$.

Solution: Since

$$(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) = \mathbf{x}\cdot\mathbf{x} + \mathbf{y}\cdot\mathbf{y} - 2\mathbf{x}\cdot\mathbf{y}$$

it follows that

$$\mathbf{x}\cdot\mathbf{y} = \frac{1}{2}\left\{|\mathbf{x}|^2 + |\mathbf{y}|^2 - |\mathbf{x}-\mathbf{y}|^2\right\}. \tag{i}$$

Since this holds for all vectors $\mathbf{x}, \mathbf{y}$ it must also hold when $\mathbf{x}$ and $\mathbf{y}$ are replaced by $\mathbf{Qx}$ and $\mathbf{Qy}$:

$$\mathbf{Qx}\cdot\mathbf{Qy} = \frac{1}{2}\left\{|\mathbf{Qx}|^2 + |\mathbf{Qy}|^2 - |\mathbf{Qx}-\mathbf{Qy}|^2\right\}.$$

By definition, an orthogonal linear transformation $\mathbf{Q}$ preserves length, i.e. $|\mathbf{Qv}| = |\mathbf{v}|$ for all vectors $\mathbf{v}$. Thus the preceding equation simplifies to

$$\mathbf{Qx}\cdot\mathbf{Qy} = \frac{1}{2}\left\{|\mathbf{x}|^2 + |\mathbf{y}|^2 - |\mathbf{x}-\mathbf{y}|^2\right\} \ . \tag{ii}$$

Since the right-hand-sides of the preceding expressions for $\mathbf{x}\cdot\mathbf{y}$ and $\mathbf{Qx}\cdot\mathbf{Qy}$ are the same, it follows that $\mathbf{Qx}\cdot\mathbf{Qy} = \mathbf{x}\cdot\mathbf{y}$.

Remark: Thus an orthogonal linear transformation preserves the length of any vector and the inner product between any two vectors. It follows therefore that an orthogonal linear transformation preserves the angle between a pair of vectors as well.

Example 2.13: Let $\mathbf{Q}$ be an orthogonal linear transformation. Show that

a. $\mathbf{Q}$ is non-singular, and that

b. $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

Solution:

a. To show that $\mathbf{Q}$ is non-singular we must show that the only vector $\mathbf{x}$ for which $\mathbf{Qx} = \mathbf{o}$ is the null vector $\mathbf{x} = \mathbf{o}$. Suppose that $\mathbf{Qx} = \mathbf{o}$ for some vector $\mathbf{x}$. Taking the norm of the two sides of this equation leads to $|\mathbf{Qx}| = |\mathbf{o}| = 0$. However an orthogonal linear transformation preserves length and therefore $|\mathbf{Qx}| = |\mathbf{x}|$. Consequently $|\mathbf{x}| = 0$. However the only vector of zero length is the null vector and so necessarily $\mathbf{x} = \mathbf{o}$. Thus $\mathbf{Q}$ is non-singular.

b. Since $\mathbf{Q}$ is orthogonal it preserves the inner product: $\mathbf{Qx}\cdot\mathbf{Qy} = \mathbf{x}\cdot\mathbf{y}$ for all vectors $\mathbf{x}$ and $\mathbf{y}$. However the property (2.19) of the transpose shows that $\mathbf{Qx}\cdot\mathbf{Qy} = \mathbf{x}\cdot\mathbf{Q}^T\mathbf{Qy}$. It follows that $\mathbf{x}\cdot\mathbf{Q}^T\mathbf{Qy} = \mathbf{x}\cdot\mathbf{y}$ for all vectors $\mathbf{x}$ and $\mathbf{y}$, and therefore that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Thus $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

Example 2.14: If α_1 and α_2 are two distinct eigenvalues of a symmetric linear transformation $\mathbf{A}$, show that the corresponding eigenvectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are orthogonal to each other.

Solution: Recall from the definition of the transpose that $\mathbf{Aa}_1\cdot\mathbf{a}_2 = \mathbf{a}_1\cdot\mathbf{A}^T\mathbf{a}_2$, and since $\mathbf{A}$ is symmetric that $\mathbf{A} = \mathbf{A}^T$. Thus

$$\mathbf{Aa}_1\cdot\mathbf{a}_2 = \mathbf{a}_1\cdot\mathbf{Aa}_2 \ .$$

Since $\mathbf{a}_1$ and $\mathbf{a}_2$ are eigenvectors of $\mathbf{A}$ corresponding to the eigenvalues α_1 and α_2, we have $\mathbf{A}\mathbf{a}_1 = \alpha_1\mathbf{a}_1$ and $\mathbf{A}\mathbf{a}_2 = \alpha_2\mathbf{a}_2$. Thus the preceding equation reduces to $\alpha_1\mathbf{a}_1 \cdot \mathbf{a}_2 = \alpha_2\mathbf{a}_1 \cdot \mathbf{a}_2$ or equivalently

$$(\alpha_1 - \alpha_2)(\mathbf{a}_1 \cdot \mathbf{a}_2) = 0.$$

Since, $\alpha_1 \neq \alpha_2$ it follows that necessarily $\mathbf{a}_1 \cdot \mathbf{a}_2 = 0$.

Example 2.15: If λ and $\mathbf{e}$ are an eigenvalue and eigenvector of an arbitrary linear transformation $\mathbf{A}$, show that λ and $\mathbf{P}^{-1}\mathbf{e}$ are an eigenvalue and eigenvector of the linear transformation $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. Here $\mathbf{P}$ is an arbitrary non-singular linear transformation.

Solution: Since $\mathbf{P}\mathbf{P}^{-1} = \mathbf{I}$ it follows that $\mathbf{A}\mathbf{e} = \mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{e}$. However we are told that $\mathbf{A}\mathbf{e} = \lambda\mathbf{e}$, whence $\mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{e} = \lambda\mathbf{e}$. Operating on both sides with $\mathbf{P}^{-1}$ gives $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{e} = \lambda\mathbf{P}^{-1}\mathbf{e}$ which establishes the result.

Example 2.16: If λ is an eigenvalue of an orthogonal linear transformation $\mathbf{Q}$, show that $|\lambda| = 1$.

Solution: Let λ and $\mathbf{e}$ be an eigenvalue and corresponding eigenvector of $\mathbf{Q}$. Thus $\mathbf{Q}\mathbf{e} = \lambda\mathbf{e}$ and so $|\mathbf{Q}\mathbf{e}| = |\lambda\mathbf{e}| = |\lambda|\,|\mathbf{e}|$. However, $\mathbf{Q}$ preserves length and so $|\mathbf{Q}\mathbf{e}| = |\mathbf{e}|$. Thus $|\lambda| = 1$.

Remark: We will show later that $+1$ is an eigenvalue of a "proper" orthogonal linear transformation on $\mathbb{E}_3$. The corresponding eigvector is known as the axis of $\mathbf{Q}$.

Example 2.17: The components of a linear transformation $\mathbf{A}$ in an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are the unique real numbers A_{ij} defined by

$$\mathbf{A}\mathbf{e}_j = \sum_{i=1}^{3} A_{ij}\mathbf{e}_i, \qquad j = 1, 2, 3. \tag{i}$$

Show that the linear transformation $\mathbf{A}$ can be represented as

$$\mathbf{A} = \sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}(\mathbf{e}_i \otimes \mathbf{e}_j). \tag{ii}$$

Solution: Consider the linear transformation given on the right-hand side of (ii) and operate it on an arbitrary vector $\mathbf{x}$:

$$\left(\sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}(\mathbf{e}_i \otimes \mathbf{e}_j)\right)\mathbf{x} = \sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}(\mathbf{x}\cdot\mathbf{e}_j)\mathbf{e}_i = \sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}x_j\mathbf{e}_i = \sum_{j=1}^{3} x_j\left(\sum_{i=1}^{3} A_{ij}\mathbf{e}_i\right),$$

where we have used the facts that $(\mathbf{p}\otimes\mathbf{q})\mathbf{r} = (\mathbf{q}\cdot\mathbf{r})\mathbf{p}$ and $x_i = \mathbf{x}\cdot\mathbf{e}_i$. On using (i) in the right most expression above, we can continue this calculation as follows:

$$\left(\sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}(\mathbf{e}_i \otimes \mathbf{e}_j)\right)\mathbf{x} = \sum_{j=1}^{3} x_j\mathbf{A}\mathbf{e}_j = \mathbf{A}\sum_{j=1}^{3} x_j\mathbf{e}_j = \mathbf{A}\mathbf{x}.$$

The desired result follows from this since this holds for arbitrary vectors $\mathbf{x}$.

Example 2.18: Let $\mathbf{R}$ be a "rotation transformation" that rotates vectors in $\mathbb{E}_3$ through an angle $\theta, 0 < \theta < \pi$, about an axis $\mathbf{e}$ (in the sense of the right-hand rule). Show that $\mathbf{R}$ can be represented as

$$\mathbf{R} = \mathbf{e} \otimes \mathbf{e} + (\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2) \cos\theta - (\mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1) \sin\theta, \qquad \text{(i)}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are any two mutually orthogonal vectors such that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}\}$ forms a right-handed orthonormal basis for $\mathbb{E}_3$.

Solution: We begin by listing what is given to us in the problem statement. Since the transformation $\mathbf{R}$ simply rotates vectors, it necessarily preserves the length of a vector and so

$$|\mathbf{Rx}| = |\mathbf{x}| \qquad \text{for all vectors } \mathbf{x}. \qquad \text{(ii)}$$

In addition, since the angle through which $\mathbf{R}$ rotates a vector is θ, the angle between any vector $\mathbf{x}$ and its image $\mathbf{Rx}$ is θ:

$$\mathbf{Rx} \cdot \mathbf{x} = |\mathbf{x}|^2 \cos\theta \qquad \text{for all vectors } \mathbf{x}. \qquad \text{(iii)}$$

Next, since $\mathbf{R}$ rotates vectors about the axis $\mathbf{e}$, the angle between any vector $\mathbf{x}$ and $\mathbf{e}$ must equal the angle between $\mathbf{Rx}$ and $\mathbf{e}$:

$$\mathbf{Rx} \cdot \mathbf{e} = \mathbf{x} \cdot \mathbf{e} \qquad \text{for all vectors } \mathbf{x}; \qquad \text{(iv)}$$

moreover, it leaves the axis $\mathbf{e}$ itself unchanged:

$$\mathbf{Re} = \mathbf{e}. \qquad \text{(v)}$$

And finally, since the rotation is in the sense of the right-hand rule, for any vector $\mathbf{x}$ that is not parallelel to the axis $\mathbf{e}$, the vectors $\mathbf{x}, \mathbf{Rx}$ and $\mathbf{e}$ must obey the inequality

$$(\mathbf{x} \times \mathbf{Rx}) \cdot \mathbf{e} > 0 \qquad \text{for all vectors } \mathbf{x} \text{ that are not parallel to } \mathbf{e}. \qquad \text{(vi)}$$

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}\}$ be a right-handed orthonormal basis. This implies that any vector in $\mathbb{E}_3$, and therefore in particular the vectors $\mathbf{Re}_1, \mathbf{Re}_2$ and $\mathbf{Re}$, can be expressed as linear combinations of $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}$,

$$\left.\begin{aligned} \mathbf{Re}_1 &= R_{11}\mathbf{e}_1 + R_{21}\mathbf{e}_2 + R_{31}\mathbf{e}, \\ \mathbf{Re}_2 &= R_{12}\mathbf{e}_1 + R_{22}\mathbf{e}_2 + R_{32}\mathbf{e}, \\ \mathbf{Re} &= R_{13}\mathbf{e}_1 + R_{23}\mathbf{e}_2 + R_{33}\mathbf{e}, \end{aligned}\right\} \qquad \text{(vii)}$$

for some unique real numbers $R_{ij}, i, j = 1, 2, 3$.

First, it follows from (v) and (vii)$_3$ that

$$R_{13} = 0, \quad R_{23} = 0, \quad R_{33} = 1.$$

Second, we conclude from (iv) with the choice $\mathbf{x} = \mathbf{e}_1$ that $\mathbf{Re}_1 \cdot \mathbf{e} = 0$. Similarly $\mathbf{Re}_2 \cdot \mathbf{e} = 0$. These together with (vii) imply that

$$R_{31} = R_{32} = 0.$$

Third, it follows from (iii) with $\mathbf{x} = \mathbf{e}_1$ and (vii)$_1$ that $R_{11} = \cos\theta$. One similarly shows that $R_{22} = \cos\theta$. Thus

$$R_{11} = R_{22} = \cos\theta.$$

Collecting these results allows us to write (vii) as

$$\left.\begin{aligned} \mathbf{Re}_1 &= \cos\theta\ \mathbf{e}_1 + R_{21}\ \mathbf{e}_2, \\ \mathbf{Re}_2 &= R_{12}\ \mathbf{e}_1 + \cos\theta\ \mathbf{e}_2, \\ \mathbf{Re} &= \mathbf{e}, \end{aligned}\right\} \tag{viii}$$

Fourth, the inequality (vi) with the choice $\mathbf{x} = \mathbf{e}_1$, together with (viii) and the fact that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}\}$ forms a right-handed basis yields $R_{21} > 0$. Similarly the choice $\mathbf{x} = \mathbf{e}_2$, yields $R_{12} < 0$. Fifth, (ii) with $\mathbf{x} = \mathbf{e}_1$ gives $|\mathbf{Re}_1| = 1$ which in view of (viii)$_1$ requires that $R_{21} = \pm\sin\theta$. Similarly we find that $R_{12} = \pm\sin\theta$. Collecting these results shows that

$$R_{21} = +\sin\theta, \qquad R_{12} = -\sin\theta,$$

since $0 < \theta < \pi$. Thus in conclusion we can write (viii) as

$$\left.\begin{aligned} \mathbf{Re}_1 &= \cos\theta\ \mathbf{e}_1 + \sin\theta\ \mathbf{e}_2, \\ \mathbf{Re}_2 &= -\sin\theta\ \mathbf{e}_1 + \cos\theta\ \mathbf{e}_2, \\ \mathbf{Re} &= \mathbf{e}. \end{aligned}\right\} \tag{ix}$$

Finally, recall the representation (2.40) of a linear transformation in terms of its components as defined in (2.39). Applying this to (ix) allows us to write

$$\mathbf{R} = \cos\theta\ (\mathbf{e}_1 \otimes \mathbf{e}_1) + \sin\theta\ (\mathbf{e}_2 \otimes \mathbf{e}_1) - \sin\theta\ (\mathbf{e}_1 \otimes \mathbf{e}_2) + \cos\theta\ (\mathbf{e}_2 \otimes \mathbf{e}_2) + (\mathbf{e} \otimes \mathbf{e}) \tag{x}$$

which can be rearranged to give the desired result.

Example 2.19: If $\mathbf{F}$ is a nonsingular linear transformation, show that $\mathbf{F}^T\mathbf{F}$ is symmetric and positive definite.

Solution: For any linear transformations $\mathbf{A}$ and $\mathbf{B}$ we know that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ and $(\mathbf{A}^T)^T = \mathbf{A}$. It therefore follows that

$$(\mathbf{F}^T\mathbf{F})^T = \mathbf{F}^T\ (\mathbf{F}^T)^T = \mathbf{F}^T\mathbf{F}; \tag{i}$$

this shows that $\mathbf{F}^T\mathbf{F}$ is symmetric.

In order to show that $\mathbf{F}^T\mathbf{F}$ is positive definite, we consider the quadratic form $\mathbf{F}^T\mathbf{F}\mathbf{x}\cdot\mathbf{x}$. By using the property (2.19) of the transpose, we can write

$$\mathbf{F}^T\mathbf{F}\mathbf{x}\cdot\mathbf{x} = (\mathbf{F}\mathbf{x})\cdot(\mathbf{F}\mathbf{x}) = |\mathbf{F}\mathbf{x}|^2 \geq 0. \tag{ii}$$

Further, equality holds here if and only if $\mathbf{F}\mathbf{x} = \mathbf{o}$, which, since $\mathbf{F}$ is nonsingular, can happen only if $\mathbf{x} = \mathbf{o}$. Thus $\mathbf{F}^T\mathbf{F}\mathbf{x}\cdot\mathbf{x} > 0$ for all vectors $\mathbf{x} \neq \mathbf{o}$ and so $\mathbf{F}^T\mathbf{F}$ is positive definite.

Example 2.20: Consider a symmetric positive definite linear transformation $\mathbf{S}$. Show that it has a unique symmetric positive definite square root, i.e. show that there is a unique symmetric positive definite linear transformation $\mathbf{T}$ for which $\mathbf{T}^2 = \mathbf{S}$.

Solution: Since $\mathbf{S}$ is symmetric and positive definite it has three real positive eigenvalues $\sigma_1, \sigma_2, \sigma_3$ with corresponding eigenvectors $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ which may be taken to be orthonormal. Further, we know that $\mathbf{S}$ can be represented as

$$\mathbf{S} = \sum_{i=1}^{3} \sigma_i (\mathbf{s}_i \otimes \mathbf{s}_i). \tag{i}$$

If one defines a linear transformation $\mathbf{T}$ by

$$\mathbf{T} = \sum_{i=1}^{3} \sqrt{\sigma_i} (\mathbf{s}_i \otimes \mathbf{s}_i) \tag{ii}$$

one can readily verify that $\mathbf{T}$ is symmetric, positive definite and that $\mathbf{T}^2 = \mathbf{S}$. This establishes the existence of a symmetric positive definite square-root of $\mathbf{S}$. What remains is to show uniqueness of this square-root.

Suppose that $\mathbf{S}$ has two symmetric positive definite square roots $\mathbf{T}_1$ and $\mathbf{T}_2$: $\mathbf{S} = \mathbf{T}_1^2 = \mathbf{T}_2^2$. Let $\sigma > 0$ and $\mathbf{s}$ be an eigenvalue and corresponding eigenvector of $\mathbf{S}$. Then $\mathbf{S}\mathbf{s} = \sigma\mathbf{s}$ and so $\mathbf{T}_1^2\mathbf{s} = \sigma\mathbf{s}$. Thus we have

$$(\mathbf{T}_1 + \sqrt{\sigma}\mathbf{I})(\mathbf{T}_1 - \sqrt{\sigma}\mathbf{I})\mathbf{s} = \mathbf{0}\ . \tag{iii}$$

If we set $\mathbf{f} = (\mathbf{T}_1 - \sqrt{\sigma}\mathbf{I})\mathbf{s}$ this can be written as

$$\mathbf{T}_1\mathbf{f} = -\sqrt{\sigma}\mathbf{f}\ . \tag{iv}$$

Thus either $\mathbf{f} = \mathbf{o}$ or $\mathbf{f}$ is an eigenvector of $\mathbf{T}_1$ corresponding to the eigenvalue $-\sqrt{\sigma}(< 0)$. Since $\mathbf{T}_1$ is positive definite it cannot have a negative eigenvalue. Thus $\mathbf{f} = \mathbf{o}$ and so

$$\mathbf{T}_1\mathbf{s} = \sqrt{\sigma}\mathbf{s}\ . \tag{v}$$

It similarly follows that $\mathbf{T}_2\mathbf{s} = \sqrt{\sigma}\mathbf{s}$ and therefore that

$$\mathbf{T}_1\mathbf{s} = \mathbf{T}_2\mathbf{s}. \tag{vi}$$

This holds for *every* eigenvector $\mathbf{s}$ of $\mathbf{S}$: i.e. $\mathbf{T}_1\mathbf{s}_i = \mathbf{T}_2\mathbf{s}_i,\ \ i = 1, 2, 3$. Since the triplet of eigenvectors form a basis for the underlying vector space this in turn implies that $\mathbf{T}_1\mathbf{x} = \mathbf{T}_2\mathbf{x}$ for any vector $\mathbf{x}$. Thus $\mathbf{T}_1 = \mathbf{T}_2$.

Example 2.21: *Polar Decomposition Theorem*: If $\mathbf{F}$ is a nonsingular linear transformation, show that there exists a unique positive definite symmetric linear transformation $\mathbf{U}$, and a unique orthogonal linear transformation $\mathbf{R}$ such that $\mathbf{F} = \mathbf{R}\mathbf{U}$.

Solution: It follows from Example 2.19 that $\mathbf{F}^T\mathbf{F}$ is symmetric and positive definite. It then follows from Example 2.20 that $\mathbf{F}^T\mathbf{F}$ has a unique symmetric positive definite square root, say, $\mathbf{U}$:

$$\mathbf{U} = \sqrt{\mathbf{F}^T\mathbf{F}}. \tag{i}$$

Finally, since $\mathbf{U}$ is positive definite, it is nonsingular, and its inverse $\mathbf{U}^{-1}$ exists. Define the linear transformation $\mathbf{R}$ through:

$$\mathbf{R} = \mathbf{F}\mathbf{U}^{-1}. \tag{ii}$$

All we have to do is to show that $\mathbf{R}$ is orthogonal. But this follows from

$$\mathbf{R}^T\mathbf{R} = (\mathbf{F}\mathbf{U}^{-1})^T\ (\mathbf{F}\mathbf{U}^{-1}) = (\mathbf{U}^{-1})^T\mathbf{F}^T\ \mathbf{F}\mathbf{U}^{-1} = \mathbf{U}^{-1}\mathbf{U}^2\mathbf{U}^{-1} = \mathbf{I}. \tag{iii}$$

In this calculation we have used the fact that $\mathbf{U}$, and so $\mathbf{U}^{-1}$, are symmetric. This establishes the proposition (except for the uniqueness which if left as an exercise).

Example 2.22: The polar decomposition theorem states that any nonsingular linear transformation $\mathbf{F}$ can be represented uniquely in the forms $\mathbf{F} = \mathbf{R}\mathbf{U} = \mathbf{V}\mathbf{R}$ where $\mathbf{R}$ is orthogonal and $\mathbf{U}$ and $\mathbf{V}$ are symmetric and positive definite. Let $\lambda_i, \mathbf{r}_i,\ i = 1, 2, 3$ be the eigenvalues and eigenvectors of $\mathbf{U}$. From Example 2.15 it follows that the eigenvalues of $\mathbf{V}$ are the same as those of $\mathbf{U}$ and that the corresponding eigenvectors $\boldsymbol{\ell}_i$ of $\mathbf{V}$ are given by $\boldsymbol{\ell}_i = \mathbf{R}\mathbf{r}_i$. Thus $\mathbf{U}$ and $\mathbf{V}$ have the spectral decompositions

$$\mathbf{U} = \sum_{i=1}^{3} \lambda_i \mathbf{r}_i \otimes \mathbf{r}_i\ , \qquad \mathbf{V} = \sum_{i=1}^{3} \lambda_i \boldsymbol{\ell}_i \otimes \boldsymbol{\ell}_i\ .$$

Show that

$$\mathbf{F} = \sum_{i=1}^{3} \lambda_i \boldsymbol{\ell}_i \otimes \mathbf{r}_i\ , \qquad \mathbf{R} = \sum_{i=1}^{3} \boldsymbol{\ell}_i \otimes \mathbf{r}_i\ .$$

Solution: First, by using the property $(2.38)_1$ and $\boldsymbol{\ell}_i = \mathbf{R}\mathbf{r}_i$ we have

$$\mathbf{F} = \mathbf{R}\mathbf{U} = \mathbf{R}\ \sum_{i=1}^{3}\ \lambda_i \mathbf{r}_i \otimes \mathbf{r}_i = \sum_{i=1}^{3}\ \lambda_i (\mathbf{R}\mathbf{r}_i) \otimes \mathbf{r}_i = \sum_{i=1}^{3}\ \lambda_i \boldsymbol{\ell}_i \otimes \mathbf{r}_i. \tag{i}$$

Next, since $\mathbf{U}$ is non-singular

$$\mathbf{U}^{-1} = \sum_{i=1}^{3}\ \lambda_i^{-1} \mathbf{r}_i \otimes \mathbf{r}_i.$$

and therefore

$$\mathbf{R} = \mathbf{F}\mathbf{U}^{-1} = \sum_{i=1}^{3}\ \lambda_i \boldsymbol{\ell}_i \otimes \mathbf{r}_i\ \sum_{j=1}^{3}\ \lambda_j^{-1} \mathbf{r}_j \otimes \mathbf{r}_j = \sum_{i=1}^{3}\sum_{j=1}^{3} \lambda_i \lambda_j^{-1} (\boldsymbol{\ell}_i \otimes \mathbf{r}_i)(\mathbf{r}_j \otimes \mathbf{r}_j).$$

By using the property $(2.37)_2$ and the fact that $\mathbf{r}_i \cdot \mathbf{r}_j = \delta_{ij}$, we have $(\boldsymbol{\ell}_i \otimes \mathbf{r}_i)(\mathbf{r}_j \otimes \mathbf{r}_j) = (\mathbf{r}_i \cdot \mathbf{r}_j)(\boldsymbol{\ell}_i \otimes \mathbf{r}_j) = \delta_{ij}(\boldsymbol{\ell}_i \otimes \mathbf{r}_j)$. Therefore

$$\mathbf{R} = \sum_{i=1}^{3}\sum_{j=1}^{3} \lambda_i \lambda_j^{-1} \delta_{ij} (\boldsymbol{\ell}_i \otimes \mathbf{r}_j) = \sum_{i=1}^{3} \lambda_i \lambda_i^{-1} (\boldsymbol{\ell}_i \otimes \mathbf{r}_i) = \sum_{i=1}^{3} (\boldsymbol{\ell}_i \otimes \mathbf{r}_i). \tag{ii}$$

Example 2.23: Determine the rank and the null space of the linear transformation $\mathbf{C} = \mathbf{a} \otimes \mathbf{b}$ where $\mathbf{a} \neq \mathbf{o}, \mathbf{b} \neq \mathbf{o}$.

Solution: Recall that the rank of any linear transformation $\mathbf{A}$ is the dimension of its range. (The range of $\mathbf{A}$ is the particular subspace of $\mathbb{E}_3$ comprised of all vectors $\mathbf{Ax}$ as $\mathbf{x}$ takes all values in $\mathbb{E}_3$.) Since $\mathbf{Cx} = (\mathbf{b}\cdot\mathbf{x})\mathbf{a}$ the vector $\mathbf{Cx}$ is parallel to the vector $\mathbf{a}$ for every choice of the vector $\mathbf{x}$. Thus the range of $\mathbf{C}$ is the set of vectors parallel to $\mathbf{a}$ and its dimension is one. The linear transformation $\mathbf{C}$ therefore has rank one.

Recall that the null space of any linear transformation $\mathbf{A}$ is the particular subspace of $\mathbb{E}_3$ comprised of the set of all vectors $\mathbf{x}$ for which $\mathbf{Ax} = \mathbf{o}$. Since $\mathbf{Cx} = (\mathbf{b}\cdot\mathbf{x})\mathbf{a}$ and $\mathbf{a} \neq \mathbf{o}$, the null space of $\mathbf{C}$ consists of all vectors $\mathbf{x}$ for which $\mathbf{b}\cdot\mathbf{x}$, i.e. the set of all vectors normal to $\mathbf{b}$.

Example 2.24: Let $\lambda_1 \leq \lambda_2 \leq \lambda_3$ be the eigenvalues of the symmetric linear transformation $\mathbf{S}$. Show that $\mathbf{S}$ can be expressed in the form

$$\mathbf{S} = (\mathbf{I} + \mathbf{a}\otimes\mathbf{b})(\mathbf{I} + \mathbf{b}\otimes\mathbf{a}) \qquad \mathbf{a} \neq \mathbf{o},\ \mathbf{b} \neq \mathbf{o}, \tag{i}$$

if and only if

$$0 \leq \lambda_1 \leq 1, \quad \lambda_2 = 1, \quad \lambda_3 \geq 1. \tag{ii}$$

Example 2.25: Calculate the square roots of the identity tensor.

Solution: The identity is certainly a symmetric positive definite tensor. By the result of a previous example on the square-root of a symmetric positive definite tensor, it follows that there is a unique symmetric positive definite tensor which is the square root of $\mathbf{I}$. Obviously, this square root is also $\mathbf{I}$. However, there are other square roots of $\mathbf{I}$ that are *not* symmetric positive definite. We are to explore them here: thus we wish to determine a tensor $\mathbf{A}$ on $\mathbb{E}_3$ such that $\mathbf{A}^2 = \mathbf{I}$, $\mathbf{A} \neq \mathbf{I}$ and $\mathbf{A} \neq -\mathbf{I}$.

First, if $\mathbf{Ax} = \mathbf{x}$ for *every* vector $\mathbf{x} \in \mathbb{E}_3$, then, by definition, $\mathbf{A} = \mathbf{I}$. Since we are given that $\mathbf{A} \neq \mathbf{I}$, there must exist at least one non-null vector $\mathbf{x}$ for which $\mathbf{Ax} \neq \mathbf{x}$; call this vector $\mathbf{f}_1$ so that $\mathbf{Af}_1 \neq \mathbf{f}_1$. Set

$$\mathbf{e}_1 = (\mathbf{A} - \mathbf{I})\,\mathbf{f}_1; \tag{i}$$

since $\mathbf{Af}_1 \neq \mathbf{f}_1$, it follows that $\mathbf{e}_1 \neq \mathbf{o}$. Observe that

$$(\mathbf{A} + \mathbf{I})\,\mathbf{e}_1 = (\mathbf{A} + \mathbf{I})\,(\mathbf{A} - \mathbf{I})\mathbf{f}_1 = (\mathbf{A}^2 - \mathbf{I})\mathbf{f}_1 = \mathbf{Of}_1 = \mathbf{o}. \tag{ii}$$

Therefore

$$\mathbf{Ae}_1 = -\mathbf{e}_1 \tag{iii}$$

and so -1 is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{e}_1$. Without loss of generality we can assume that $|\mathbf{e}_1| = 1$.

Second, the fact that $\mathbf{A} \neq -\mathbf{I}$, together with $\mathbf{A}^2 = \mathbf{I}$ similary implies that there must exist a unit vector $\mathbf{e}_2$ for which

$$\mathbf{Ae}_2 = \mathbf{e}_2, \tag{iv}$$

from which we conclude that $+1$ is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{e}_2$.

Third, one can show that $\{\mathbf{e}_1, \mathbf{e}_2\}$ is a linearly independent pair of vectors. To see this, suppose that for some scalars ξ_1, ξ_2 one has

$$\xi_1 \mathbf{e}_1 + \xi_2 \mathbf{e}_2 = \mathbf{o}.$$

Operating on this by $\mathbf{A}$ yields $\xi_1 \mathbf{A}\mathbf{e}_1 + \xi_2 \mathbf{A}\mathbf{e}_2 = \mathbf{o}$, which on using (iii) and (iv) leads to

$$-\xi_1 \mathbf{e}_1 + \xi_2 \mathbf{e}_2 = \mathbf{o}.$$

Subtracting and adding the preceding two equations shows that $\xi_1 \mathbf{e}_1 = \xi_2 \mathbf{e}_2 = \mathbf{o}$. Since $\mathbf{e}_1$ and $\mathbf{e}_2$ are eigenvectors, neither of them is the null vector $\mathbf{o}$, and therefore $\xi_1 = \xi_2 = 0$. Therefore $\mathbf{e}_1$ and $\mathbf{e}_2$ are linearly independent.

Fourth, let $\mathbf{e}_3$ be a unit vector that is perpendicular to both $\mathbf{e}_1$ and $\mathbf{e}_2$. The triplet of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is linearly independent and therefore forms a basis for $\mathbb{E}_3$.

Fifth, the components A_{ij} of the tensor $\mathbf{A}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are given, as usual, by

$$\mathbf{A}\mathbf{e}_j = A_{ij}\mathbf{e}_i. \tag{v}$$

Comparing (v) with (iii) yields $A_{11} = -1, A_{21} = A_{31} = 0$, and similarly comparing (v) with (iv) yields $A_{22} = 1, A_{12} = A_{32} = 0$. The matrix of components of $\mathbf{A}$ in this basis is therefore

$$[A] = \begin{pmatrix} -1 & 0 & A_{13} \\ 0 & 1 & A_{23} \\ 0 & 0 & A_{33} \end{pmatrix}. \tag{vi}$$

It follows that

$$[A^2] = [A]^2 = [A][A] = \begin{pmatrix} 1 & 0 & -A_{13} + A_{13}A_{33} \\ 0 & 1 & A_{23} + A_{23}A_{33} \\ 0 & 0 & A_{33}^2 \end{pmatrix}. \tag{vii}$$

(Notation: $[A^2]$ is the matrix of components of $\mathbf{A}^2$ while $[A]^2$ is the square of the matrix of components of $\mathbf{A}$. Why is $[A^2] = [A]^2$?) However, since $\mathbf{A}^2 = \mathbf{I}$, the matrix of components of $\mathbf{A}^2$ in *any* basis has to be the identity matrix. Therefore we must have

$$-A_{13} + A_{13}A_{33} = 0, \qquad A_{23} + A_{23}A_{33} = 0, \qquad A_{33}^2 = 1, \tag{viii}$$

which implies that

$$\text{either} \quad \left.\begin{array}{rcl} A_{13} & = & \text{arbitrary}, \\ A_{23} & = & 0, \\ A_{33} & = & 1, \end{array}\right\} \qquad \text{or} \quad \left.\begin{array}{rcl} A_{13} & = & 0, \\ A_{23} & = & \text{arbitrary}, \\ A_{33} & = & -1. \end{array}\right\} \tag{ix}$$

Consequently the matrix $[A]$ must necessarily have one of the two forms

$$\begin{pmatrix} -1 & 0 & \alpha_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & \alpha_2 \\ 0 & 0 & -1 \end{pmatrix}, \tag{x}$$

where α_1 and α_2 are arbitrary scalars.

Sixth, set

$$\mathbf{p}_1 = \mathbf{e}_1, \qquad \mathbf{q}_1 = -\mathbf{e}_1 + \frac{\alpha_1}{2}\,\mathbf{e}_3. \tag{xi}$$

Then

$$\mathbf{p}_1 \otimes \mathbf{q}_1 = -\mathbf{e}_1 \otimes \mathbf{e}_1 + \frac{\alpha_1}{2}\mathbf{e}_1 \otimes \mathbf{e}_3,$$

and therefore

$$\begin{aligned} \mathbf{I} + 2\mathbf{p}_1 \otimes \mathbf{q}_1 &= \Big(\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3\Big) - 2\mathbf{e}_1 \otimes \mathbf{e}_1 + \alpha_1 \mathbf{e}_1 \otimes \mathbf{e}_3 \\ &= -\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3 + \alpha_1 \mathbf{e}_1 \otimes \mathbf{e}_3. \end{aligned}$$

Note from this that the components of the tensor $\mathbf{I} + 2\mathbf{p}_1 \otimes \mathbf{q}_1$ are given by $(\mathrm{x})_1$. Conversely, one can readily verify that the tensor

$$\mathbf{A} = \mathbf{I} + 2\mathbf{p}_1 \otimes \mathbf{q}_1 \tag{xii}$$

has the desired properties $\mathbf{A}^2 = \mathbf{I}, \mathbf{A} \neq \mathbf{I}, \mathbf{A} \neq -\mathbf{I}$ for any value of the scalar α_1.

Alternatively set

$$\mathbf{p}_2 = \mathbf{e}_2, \qquad \mathbf{q}_2 = \mathbf{e}_2 + \frac{\alpha_2}{2}\,\mathbf{e}_3. \tag{xiii}$$

Then

$$\mathbf{p}_2 \otimes \mathbf{q}_2 \;=\; \mathbf{e}_2 \otimes \mathbf{e}_2 + \frac{\alpha_2}{2}\mathbf{e}_2 \otimes \mathbf{e}_3,$$

and therefore

$$\begin{aligned} -\mathbf{I} + 2\mathbf{p}_2 \otimes \mathbf{q}_2 &= \Big(-\mathbf{e}_1 \otimes \mathbf{e}_1 - \mathbf{e}_2 \otimes \mathbf{e}_2 - \mathbf{e}_3 \otimes \mathbf{e}_3\Big) + 2\mathbf{e}_2 \otimes \mathbf{e}_2 + \alpha_2 \mathbf{e}_2 \otimes \mathbf{e}_3 \\ &= -\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 - \mathbf{e}_3 \otimes \mathbf{e}_3 + \alpha_2 \mathbf{e}_2 \otimes \mathbf{e}_3. \end{aligned}$$

Note from this that the components of the tensor $-\mathbf{I} + 2\mathbf{p}_2 \otimes \mathbf{q}_2$ are given by $(\mathrm{x})_2$. Conversely, one can readily verify that the tensor

$$\mathbf{A} = -\mathbf{I} + 2\mathbf{p}_2 \otimes \mathbf{q}_2 \tag{xiv}$$

has the desired properties $\mathbf{A}^2 = \mathbf{I}, \mathbf{A} \neq \mathbf{I}, \mathbf{A} \neq -\mathbf{I}$ for any value of the scalar α_2.

Thus the tensors defined in (xii) and (xiv) are both square roots of the identity tensor that are not symmetric positive definite.

References

1. I.M. Gelfand, *Lectures on Linear Algebra*, Wiley, New York, 1963.
2. P.R. Halmos, *Finite Dimensional Vector Spaces*, Van Nostrand, New Jersey, 1958.
3. J.K. Knowles, *Linear Vector Spaces and Cartesian Tensors*, Oxford University Press, New York, 1997.

Chapter 3

Components of Vectors and Tensors. Cartesian Tensors.

Notation:

α	.....	scalar
$\{a\}$	.....	3×1 column matrix
$\mathbf{a}$	.....	vector
a_i	.....	i^{th} component of the vector $\mathbf{a}$ in some basis; or i^{th} element of the column matrix $\{a\}$
$[A]$	.....	3×3 square matrix
$\mathbf{A}$	.....	linear transformation
A_{ij}	.....	i, j component of the linear transformation $\mathbf{A}$ in some basis; or i, j element of the square matrix $[A]$
$\mathbb{C}_{ijk\ell}$	.....	i, j, k, ℓ component of 4-tensor $\mathbb{C}$ in some basis
$\mathbb{T}_{i_1 i_2 \dots i_n}$	.....	$i_1 i_2 \dots i_n$ component of n-tensor $\mathbb{T}$ in some basis.

3.1 Components of a vector in a basis.

Let $\mathbb{E}_3$ be a three-dimensional Euclidean vector space. A set of three linearly independent vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms a **basis** for $\mathbb{E}_3$ in the sense that an arbitrary vector $\mathbf{v}$ can always be expressed as a linear combination of the three basis vectors; i.e. given any $\mathbf{v} \in \mathbb{E}_3$, there are unique scalars α, β, γ such that

$$\mathbf{v} = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2 + \gamma \mathbf{e}_3. \tag{3.1}$$

If each basis vector $\mathbf{e}_i$ has unit length, and if each pair of basis vectors $\mathbf{e}_i, \mathbf{e}_j$ are mutually orthogonal, we say that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms an **orthonormal basis** for $\mathbb{E}_3$. Thus, for an orthonormal basis,

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \tag{3.2}$$

where δ_{ij} is the Kronecker delta. In these notes we shall always restrict attention to orthonormal bases unless explicitly stated otherwise. If the basis is right-handed, one has in addition that

$$\mathbf{e}_i \cdot (\mathbf{e}_j \times \mathbf{e}_k) = e_{ijk} \tag{3.3}$$

where e_{ijk} is the alternator introduced previously in (1.44).

The **components** v_i of a vector $\mathbf{v}$ in a basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are defined by

$$v_i = \mathbf{v} \cdot \mathbf{e}_i. \tag{3.4}$$

The vector can be expressed in terms of its components and the basis vectors as

$$\mathbf{v} = v_i \mathbf{e}_i. \tag{3.5}$$

The components of $\mathbf{v}$ may be assembled into a column matrix

$$\{v\} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \tag{3.6}$$

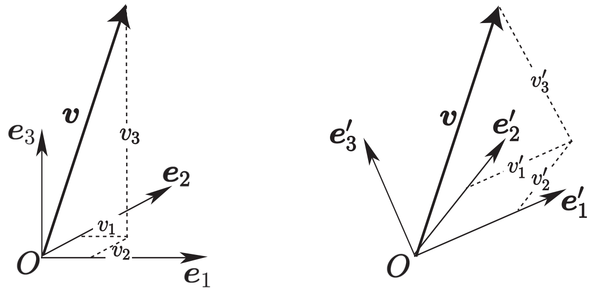


Figure 3.1: Components $\{v_1, v_2, v_3\}$ and $\{v_1', v_2', v_3'\}$ of the same vector $\mathbf{v}$ in two different bases.

Even though this is obvious from the definition (3.4), it is still important to emphasize that the components v_i of a vector depend on both the vector $\mathbf{v}$ *and* the choice of basis. Suppose, for example, that we are given two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ as shown in

Figure 3.1. Then the vector $\mathbf{v}$ has one set of components v_i in the first basis and a different set of components v_i' in the second basis:

$$v_i = \mathbf{v} \cdot \mathbf{e}_i, \qquad v_i' = \mathbf{v} \cdot \mathbf{e}_i' \ . \tag{3.7}$$

Thus the one vector $\mathbf{v}$ can be expressed in either of the two equivalent forms

$$\mathbf{v} = v_i \mathbf{e}_i \qquad \text{or} \qquad \mathbf{v} = v_i' \mathbf{e}_i'. \tag{3.8}$$

The components v_i and v_i' are related to each other (as we shall discuss later) but in general, $v_i \neq v_i'$.

Once a basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is chosen and fixed, there is a unique vector $\mathbf{x}$ associated with any given column matrix $\{x\}$ such that the components of $\mathbf{x}$ in $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are $\{x\}$. Thus, *once* the basis is fixed, there is a one-to-one correspondence between column matrices and vectors. It follows, for example, that once the basis is fixed, the vector equation $\mathbf{z} = \mathbf{x} + \mathbf{y}$ can be written equivalently as

$$\{z\} = \{x\} + \{y\} \quad \text{or} \quad z_i = x_i + y_i \tag{3.9}$$

in terms of the components x_i, y_i and z_i in the given basis.

If u_i and v_i are the components of two vectors $\mathbf{u}$ and $\mathbf{v}$ in a basis, then the scalar-product $\mathbf{u} \cdot \mathbf{v}$ can be expressed as

$$\mathbf{u} \cdot \mathbf{v} = u_i v_i; \tag{3.10}$$

the vector-product $\mathbf{u} \times \mathbf{v}$ can be expressed as

$$\mathbf{u} \times \mathbf{v} = (e_{ijk} u_j v_k) \mathbf{e}_i \quad \text{or equivalently as} \quad (\mathbf{u} \times \mathbf{v})_i = e_{ijk} u_j v_k \ , \tag{3.11}$$

where e_{ijk} is the alternator introduced previously in (1.44).

3.2 Components of a linear transformation in a basis.

Consider a linear transformation $\mathbf{A}$. Any vector in $\mathbb{E}_3$ can be expressed as a linear combination of the basis vectors $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}_3$. In particular this is true of the three vectors $\mathbf{A}\mathbf{e}_1, \mathbf{A}\mathbf{e}_2$ and $\mathbf{A}\mathbf{e}_3$. Let A_{ij} be the *ith* component of the vector $\mathbf{A}\mathbf{e}_j$ so that

$$\mathbf{A}\mathbf{e}_j = A_{ij} \ \mathbf{e}_i. \tag{3.12}$$

We can also write

$$A_{ij} = \mathbf{e}_i \cdot (\mathbf{A}\mathbf{e}_j). \tag{3.13}$$

The 9 scalars A_{ij} are known as **the components of the linear transformation A** in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. The components A_{ij} can be assembled into a square matrix:

$$[A] = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}. \tag{3.14}$$

The linear transformation $\mathbf{A}$ can be expressed in terms of its components A_{ij} and the basis vectors $\mathbf{e}_i$ as

$$\mathbf{A} = \sum_{j=1}^{3} \sum_{i=1}^{3} A_{ij}(\mathbf{e}_i \otimes \mathbf{e}_j). \tag{3.15}$$

The components A_{ij} of a linear transformation depend on both the linear transformation $\mathbf{A}$ *and* the choice of basis. Suppose, for example, that we are given two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Then the linear transformation $\mathbf{A}$ has one set of components A_{ij} in the first basis and a different set of components A'_{ij} in the second basis:

$$A_{ij} = \mathbf{e}_i \cdot (\mathbf{A}\mathbf{e}_j), \qquad A'_{ij} = \mathbf{e}'_i \cdot (\mathbf{A}\mathbf{e}'_j). \tag{3.16}$$

The components A_{ij} and A'_{ij} are related to each other (as we shall discuss later) but in general $A_{ij} \neq A'_{ij}$.

The components of the linear transformation $\mathbf{A} = \mathbf{a} \otimes \mathbf{b}$ are

$$A_{ij} = a_i b_j. \tag{3.17}$$

Once a basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is chosen and fixed, there is a unique linear transformation $\mathbf{M}$ associated with any given square matrix $[M]$ such that the components of $\mathbf{M}$ in $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are $[M]$. Thus, *once* the basis is fixed, there is a one-to-one correspondence between square matrices and linear transformations. It follows, for example, that the equation $\mathbf{y} = \mathbf{A}\mathbf{x}$ relating the linear transformation $\mathbf{A}$ and the vectors $\mathbf{x}$ and $\mathbf{y}$ can be written equivalently as

$$\{y\} = [A]\{x\} \quad \text{or} \quad y_i = \ A_{ij}x_j \tag{3.18}$$

in terms of the components A_{ij}, x_i and y_i in the given basis. Similarly, if $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ are linear transformations such that $\mathbf{C} = \mathbf{A}\mathbf{B}$, then their component matrices $[A], [B]$ and $[C]$ are related by

$$[C] = [A][B] \quad \text{or} \quad C_{ij} = \ A_{ik}B_{kj}. \tag{3.19}$$

The component matrix $[I]$ of the identity linear transformation $\mathbf{I}$ in any orthonormal basis is the unit matrix; its components are therefore given by the Kronecker delta δ_{ij}. If $[A]$ and $[A^T]$ are the component matrices of the linear transformations $\mathbf{A}$ and $\mathbf{A}^T$, then $[A^T] = [A]^T$ and $A^T_{ij} = A_{ji}$.

As mentioned in Section 2.2, a symmetric linear transformation $\mathbf{S}$ has three real eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and corresponding orthonormal eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. The eigenvectors are referred to as the *principal directions* of $\mathbf{S}$. The particular basis consisting of the eigenvectors is called a *principal basis* for $\mathbf{S}$. The component matrix $[S]$ of the symmetric linear transformation $\mathbf{S}$ in its principal basis is

$$[S] = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \tag{3.20}$$

As a final remark we note that if we are to establish certain results for vectors and linear transformations, we can, if it is more convenient to do so, pick and fix a basis, and then work with the components in that basis. If necessary, we can revert back to the vectors and linear transformations at the end. For example the first example in the previous chapter asked us to show that $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$. In terms of components, the left hand side of this reads $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_i(\mathbf{b} \times \mathbf{c})_i = a_i e_{ijk} b_j c_k = e_{ijk} a_i b_j c_k$. Similarly the right-hand side reads $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = b_i(\mathbf{c} \times \mathbf{a})_i = b_i e_{ijk} c_j a_k = e_{ijk} a_k b_i c_j$. Since i, j, k are dummy subscripts in the right-most expression, they can be changed to any other subscript; thus by changing $k \to i, i \to j$ and $j \to k$ we can write $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = e_{jki} a_i b_j c_k$. Finally recalling that the sign of e_{ijk} changes when any two adjacent subscripts are switched we find that $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = e_{jki} a_i b_j c_k = -e_{jik} a_i b_j c_k = e_{ijk} a_i b_j c_k$ where we have first switched the ki and then the ji in the subscript of the alternator. The right-most expressions of $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ and $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$ are identical and therefore this establishes the desired identity.

3.3 Components in two bases.

Consider a 3-dimensional Euclidean vector space together with two orthonormal bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Since $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms a basis, any vector, and therefore in particular the vectors $\mathbf{e}'_i$, can be represented as a linear combination of the basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let Q_{ij} be the jth component of the vector $\mathbf{e}'_i$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$:

$$\mathbf{e}'_i = Q_{ij}\mathbf{e}_j. \tag{3.21}$$

By taking the dot-product of (NNN) with $\mathbf{e}'_k$, one sees that

$$Q_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j, \tag{3.22}$$

and so Q_{ij} is the cosine of the angle between the basis vectors $\mathbf{e}'_i$ and $\mathbf{e}_j$. Observe from (NNN) that Q_{ji} can also be interpreted as the jth component of $\mathbf{e}_i$ in the basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ whence we also have

$$\mathbf{e}_i = Q_{ji}\ \mathbf{e}'_j. \tag{3.23}$$

The 9 numbers Q_{ij} can be assembled into a square matrix $[Q]$. This matrix relates the two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Since both bases are orthonormal it can be readily shown that $[Q]$ is an orthogonal matrix. If in addition, if one basis can be rotated into the other, which means that both bases are right-handed or both are left-handed, then $[Q]$ is a proper orthogonal matrix and $\det[Q] = +1$; if the two bases are related by a reflection, which means that one basis is right-handed and the other is left-handed, then $[Q]$ is an improper orthogonal matrix and $\det[Q] = -1$.

We may now **relate the different components of a single vector v** in two bases. Let v_i and v'_i be the ith component of the same vector $\mathbf{v}$ in the two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Then one can show that

$$v'_i = Q_{ij} v_j \qquad \text{or equivalently} \qquad \{v'\} = [Q]\{v\} \tag{3.24}$$

Since $[Q]$ is orthogonal, one also has the inverse relationships

$$v_i = Q_{ji} v'_j \qquad \text{or equivalently} \qquad \{v\} = [Q]^T\{v'\}. \tag{3.25}$$

In general, the component matrices $\{v\}$ and $\{v'\}$ of a vector $\mathbf{v}$ in two different bases are different. A vector whose components in every basis happen to be the same is called an **isotropic vector**: $\{v\} = [Q]\{v\}$ for all orthogonal matrices $[Q]$. It is possible to show that the only isotropic vector is the null vector $\mathbf{o}$.

Similarly, we may **relate the different components of a single linear transformation A** in two bases. Let A_{ij} and A'_{ij} be the ij-components of the same linear transformation $\mathbf{A}$ in the two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Then one can show that

$$A'_{ij} = Q_{ip} Q_{jq} A_{pq} \qquad \text{or equivalently} \qquad [A'] = [Q][A][Q]^T. \tag{3.26}$$

Since $[Q]$ is orthogonal, one also has the inverse relationships

$$A_{ij} = Q_{pi} Q_{qj} A'_{pq} \qquad \text{or equivalently} \qquad [A] = [Q]^T[A'][Q]. \tag{3.27}$$

In general, the component matrices $[A]$ and $[A']$ of a linear transformation $\mathbf{A}$ in two different bases are different. A linear transformation whose components in every basis happen to be the same is called an **isotropic linear transformation**: $[A] = [Q][A][Q]^T$ for all orthogonal matrices $[Q]$. It is possible to show that the most general isotropic symmetric linear transformation is a scalar multiple of the identity $\alpha\mathbf{I}$ where α is an arbitrary scalar.

3.4 Scalar-valued functions of linear transformations. Determinant, trace, scalar-product and norm.

Let $\Phi(\mathbf{A}; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ be a scalar-valued function that depends on a linear transformation $\mathbf{A}$ and a (non-necessarily orthonormal) basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. For example $\Phi(\mathbf{A}; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \mathbf{A}\mathbf{e}_1 \cdot \mathbf{e}_1$. Certain such functions are in fact independent of the basis, so that for every two (not-necessarily orthonormal) bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ one has $\Phi(\mathbf{A}; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \Phi(\mathbf{A}; \mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)$, and in such a case we can simply write $\Phi(\mathbf{A})$. One example of such a function is

$$\Phi(\mathbf{A}; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \frac{(\mathbf{A}\mathbf{e}_1 \times \mathbf{A}\mathbf{e}_2) \cdot \mathbf{A}\mathbf{e}_3}{(\mathbf{e}_1 \times \mathbf{e}_2) \cdot \mathbf{e}_3}, \tag{3.28}$$

(though it is certainly not obvious that this function is independent of the choice of basis).

Equivalently, let $\mathbf{A}$ be a linear transformation and let $[A]$ be the components of $\mathbf{A}$ in some basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Let $\phi([A])$ be some real-valued function defined on the set of all square matrices. If $[A']$ are the components of $\mathbf{A}$ in some other basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$, then in general $\phi([A]) \neq \phi[A'])$. This means that the function ϕ depends on the linear transformation $\mathbf{A}$ *and* the underlying basis. Certain functions ϕ have the property that $\phi([A]) = \phi[A'])$ for all pairs of bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ and therefore such a function depends on the linear transformation only and not the basis. For such a function we may write $\phi(\mathbf{A})$.

We first consider two important examples here. Since the components $[A]$ and $[A']$ of a linear tranformation $\mathbf{A}$ in two bases are related by $[A'] = [Q][A][Q]^T$, if we take the determinant of this matrix equation we get

$$\det[A'] = \det([Q][A][Q]^T) = \det[Q]\det[A]\det[Q]^T = (\det[Q])^2\det[A] = \det[A], \tag{3.29}$$

since the determinant of an orthogonal matrix is ±1. Therefore without ambiguity we may define the **determinant** of a linear transformation $\mathbf{A}$ to be the (basis independent) scalar-valued function given by

$$\det\mathbf{A} = \det[A]. \tag{3.30}$$

We will see in an example at the end of this Chapter that the particular function Φ defined in (3.28) is in fact the determinant $\det \mathbf{A}$.

Similarly, we may define the **trace** of a linear transformation $\mathbf{A}$ to be the (basis independent) scalar-valued function given by

$$\text{trace } \mathbf{A} = \text{tr}[A]. \tag{3.31}$$

In terms of its components in a basis one has

$$\det \mathbf{A} = e_{ijk} A_{1i} A_{2j} A_{3k} = e_{ijk} A_{i1} A_{j2} A_{k3}, \qquad \text{trace}\mathbf{A} = A_{ii}; \tag{3.32}$$

see (1.46). It is useful to note the following properties of the determinant of a linear transformation:

$$\det(\mathbf{AB}) = \det(\mathbf{A})\ \det(\mathbf{B}), \quad \det(\alpha\mathbf{A}) = \alpha^3\ \det(\mathbf{A}), \quad \det(\mathbf{A}^T) = \det\ (\mathbf{A}). \tag{3.33}$$

As mentioned previously, a linear transformation $\mathbf{A}$ is said to be non-singular if the only vector $\mathbf{x}$ for which $\mathbf{Ax} = \mathbf{o}$ is the null vector $\mathbf{x} = \mathbf{o}$. Equivalently, one can show that $\mathbf{A}$ is non-singular if and only if

$$\det \mathbf{A} \neq 0. \tag{3.34}$$

If $\mathbf{A}$ is non-singular, then

$$\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A}). \tag{3.35}$$

Suppose that λ and $\mathbf{v} \neq \mathbf{o}$ are an eigenvalue and eigenvector of given a linear transformation $\mathbf{A}$. Then by definition, $\mathbf{Av} = \lambda\mathbf{v}$, or equivalently $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{o}$. Since $\mathbf{v} \neq \mathbf{o}$ it follows that $\mathbf{A} - \lambda\mathbf{I}$ must be singular and so

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{3.36}$$

The eigenvalues are the roots λ of this cubic equation. The eigenvalues and eigenvectors of a linear transformation do not depend on any choice of basis. Thus the eigenvalues of a linear transformation are also scalar-valued functions of $\mathbf{A}$ whose values depends only on $\mathbf{A}$ and not the basis: $\lambda_i = \lambda_i(\mathbf{A})$. If $\mathbf{S}$ is symmetric, its matrix of components *in a principal basis* are

$$[S] = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \tag{3.37}$$

The particular scalar-valued functions

$$\begin{aligned} I_1(\mathbf{A}) &= \text{tr } \mathbf{A}, \\ I_2(\mathbf{A}) &= 1/2\left[(\text{tr } \mathbf{A})^2 - \text{tr } (\mathbf{A}^2)\right], \\ I_3(\mathbf{A}) &= \det \mathbf{A}, \end{aligned} \tag{3.38}$$

will appear frequently in what follows. It can be readily verified that for any linear transformation $\mathbf{A}$ and all orthogonal linear transformations $\mathbf{Q}$,

$$I_1(\mathbf{Q}^T\mathbf{A}\mathbf{Q}) = I_1(\mathbf{A}), \qquad I_2(\mathbf{Q}^T\mathbf{A}\mathbf{Q}) = I_2(\mathbf{A}), \qquad I_3(\mathbf{Q}^T\mathbf{A}\mathbf{Q}) = I_3(\mathbf{A}), \tag{3.39}$$

and for this reason the three functions (3.38) are said to be invariant under orthogonal transformations. Observe from (3.37) that for a symmetric linear transformation with eigenvalues $\lambda_1, \lambda_2, \lambda_3$

$$\begin{aligned} I_1(\mathbf{S}) &= \lambda_1 + \lambda_2 + \lambda_3, \\ I_2(\mathbf{S}) &= \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1, \\ I_3(\mathbf{S}) &= \lambda_1\lambda_2\lambda_3. \end{aligned} \tag{3.40}$$

The mapping (3.40) between invariants and eigenvalues is one-to-one. In addition one can show that for any linear transformation $\mathbf{A}$ and *any* real number α,

$$\det(\mathbf{A} - \alpha\mathbf{I}) = -\alpha^3 + I_1(\mathbf{A})\alpha^2 - I_2(\mathbf{A})\alpha + I_3(\mathbf{A}).$$

Note in particular that the cubic equation for the eigenvalues of a linear transformation can be written as

$$\lambda^3 - I_1(\mathbf{A})\lambda^2 + I_2(\mathbf{A})\lambda - I_3(\mathbf{A}) = 0.$$

Finally, one can show that

$$\mathbf{A}^3 - I_1(\mathbf{A})\mathbf{A}^2 + I_2(\mathbf{A})\mathbf{A} - I_3(\mathbf{A})\mathbf{I} = \mathbf{O}. \tag{3.41}$$

which is known as the Cayley-Hamilton theorem.

One can similarly define scalar-valued functions of two linear transformations $\mathbf{A}$ and $\mathbf{B}$. The particular function $\phi(\mathbf{A}, \mathbf{B})$ defined by

$$\phi(\mathbf{A}, \mathbf{B}) = \text{tr}(\mathbf{A}\mathbf{B}^T) \tag{3.42}$$

will play an important role in what follows. Note that in terms of components in a basis,

$$\phi(\mathbf{A}, \mathbf{B}) = \text{tr}(\mathbf{A}\mathbf{B}^T) = A_{ij}B_{ij}. \tag{3.43}$$

This particular scalar-valued function is often known as **the scalar product of the two linear transformation $\mathbf{A}$ and $\mathbf{B}$** and is written as $\mathbf{A} \cdot \mathbf{B}$:

$$\mathbf{A} \cdot \mathbf{B} = \mathrm{tr}(\mathbf{A}\mathbf{B}^T). \tag{3.44}$$

It is natural then to define the **magnitude** (or norm) of a linear transformation $\mathbf{A}$, denoted by $|\mathbf{A}|$ as

$$|\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}} = \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^T)}. \tag{3.45}$$

Note that in terms of components in a basis,

$$|\mathbf{A}|^2 = A_{ij}A_{ij}. \tag{3.46}$$

Observe the useful property that if $|\mathbf{A}| \to 0$, then each component

$$A_{ij} \to 0. \tag{3.47}$$

This will be used later when we linearize the theory of large deformations.

3.5 Cartesian Tensors

Consider two orthonormal bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. A quantity whose components v_i and v'_i in these two bases are related by

$$v'_i = Q_{ij}v_j \tag{3.48}$$

is called a 1st-*order Cartesian tensor* or a 1-tensor. It follows from our preceding discussion that a vector is a 1-tensor.

A quantity whose components A_{ij} and A'_{ij} in two bases are related by

$$A'_{ij} = Q_{ip}Q_{jq}A_{pq} \tag{3.49}$$

is called a 2nd-*order Cartesian tensor* or a 2-tensor. It follows from our preceding discussion that a linear transformation is a 2-tensor.

The concept of an nth-*order tensor* can be introduced similarly: let $\mathbb{T}$ be a physical entity which, in a given basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, is defined completely by a set of 3^n ordered numbers $\mathbb{T}_{i_1 i_2 \ldots i_n}$. The numbers $\mathbb{T}_{i_1 i_2 \ldots i_n}$ are called the *components* of $\mathbb{T}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. If, for example, $\mathbb{T}$ is a scalar, vector or linear transformation, it is represented by $3^0, 3^1$ and 3^2

components respectively in the given basis. Let $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ be a second basis related to the first one by the orthogonal matrix [Q], and let $\mathbb{T}'_{i_1 i_2 .... i_n}$ be the components of the entity $\mathbb{T}$ in the second basis. Then, if for every pair of such bases, these two sets of components are related by

$$\mathbb{T}'_{i_1 i_2 .... i_n} \; = \; Q_{i_1 j_1} \; Q_{i_2 j_2} \; .... \; Q_{i_n j_n} \; \mathbb{T}_{j_1 j_2 .... j_n}, \tag{3.50}$$

the entity $\mathbb{T}$ is called a n^{th}-**order Cartesian tensor** or more simply an **n-tensor**.

Note that the components of a tensor in an arbitrary basis can be calculated if its components in any one basis are known.

Two tensors of the same order are *added* by adding corresponding components.

Recall that the outer-product of two vectors **a** and **b** is the 2−tensor $\mathbf{C} = \mathbf{a} \otimes \mathbf{b}$ whose components are given by $C_{ij} = a_i \, b_j$. This can be generalized to higher-order tensors. Given an n-tensor $\mathbb{A}$ and an m-tensor $\mathbb{B}$ their **outer-product** is the $(m+n)-$tensor $\mathbb{C} = \mathbb{A} \otimes \mathbb{B}$ whose components are given by

$$\mathbb{C}_{i_1 i_2 .. i_n j_1 j_2 .. j_m} = \mathbb{A}_{i_1 i_2 ... i_n} \, \mathbb{B}_{j_1 j_2 ... j_m}. \tag{3.51}$$

Let **A** be a 2-tensor with components A_{ij} in some basis. Then "*contracting*" **A** over its subscripts leads to the scalar A_{ii}. This can be generalized to higher-order tensors. Let $\mathbb{A}$ be a n-tensor with components $\mathbb{A}_{i_1 i_2 ... i_n}$ in some basis. Then "*contracting*" $\mathbb{A}$ over two of its subscripts, say the $i_j th$ and $i_k th$ subscripts, leads to the $(n-2)-$tensor whose components in this basis are $\mathbb{A}_{i_1 \; i_2 \; .. \; i_{j-1} \; p \; i_{j+1} \; ... i_{k-1} \; p \; i_{k+1} \; .... \; i_n}$. Contracting over two subscripts involves setting those two subscripts equal, and therefore summing over them.

Let $\mathbf{a}, \mathbf{b}$ and **T** be entities whose components in a basis are denoted by a_i, b_i and T_{ij}. Suppose that the components of **T** in some basis are related to the components of **a** and **b** in that same basis by $a_i = T_{ij} b_j$. If **a** and **b** are 1-tensors, then one can readily show that **T** is necessarily a 2-tensor. This is called the **quotient rule** since it has the appearance of saying that the quotient of two 1-tensors is a 2-tensor. This rule generalizes naturally to tensors of more general order. Suppose that $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{T}$ are entities whose components in a basis are related by,

$$\mathbb{A}_{i_1 i_2 .. i_n} = \mathbb{T}_{k_1 k_2 ... k_\ell} \; \mathbb{B}_{j_1 j_2 ... j_m} \tag{3.52}$$

where some of the subscripts maybe repeated. If it is known that $\mathbb{A}$ and $\mathbb{B}$ are tensors, then $\mathbb{T}$ is necessarily a tensor as well.

In general, the components of a tensor $\mathbb{T}$ in two different bases are different: $\mathbb{T}'_{i_1 i_2 ... i_n} \neq \mathbb{T}_{i_1 i_2 ... i_n}$. However, there are certain special tensors whose components in one basis are the same as those in *any* other basis; an example of this is the identity 2-tensor **I**. Such a tensor is said to be isotropic. In general, a tensor $\mathbb{T}$ is said to be an **isotropic tensor** if its components have the same values in all bases, i.e. if

$$\mathbb{T}'_{i_1 i_2 \ldots i_n} = \mathbb{T}_{i_1 i_2 \ldots i_n} \tag{3.53}$$

in all bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Equivalently, for an isotropic tensor

$$\mathbb{T}_{i_1 i_2 \ldots . i_n} = Q_{i_1 j_1}\ Q_{i_2 j_2}\ \ldots . Q_{i_n j_n}\ \mathbb{T}_{j_1 j_2 \ldots . j_n} \quad \text{for all orthogonal matrices } [Q]. \tag{3.54}$$

One can show that (a) the only isotropic 1-tensor is the null vector $\boldsymbol{o}$; (b) the most general isotropic 2-tensor is a scalar multiple of the identity linear transformation, $\alpha\mathbf{I}$; (c) the most general isotropic 3-tensor is the null 3-tensor $\boldsymbol{o}$; (d) and the most general isotropic 4-tensor $\mathbb{C}$ has components (in any basis)

$$\mathbb{C}_{ijkl} = \alpha\delta_{ij}\delta_{kl} + \beta\delta_{ik}\delta_{jl} + \gamma\delta_{il}\delta_{jk} \tag{3.55}$$

where α, β, γ are arbitrary scalars.

3.6 Worked Examples.

In some of the examples below, we are asked to establish certain results for vectors and linear transformations. As noted previously, whenever it is more convenient we may pick and fix a basis, and then work using components in that basis. If necessary, we can revert back to the vectors and linear transformations at the end. We shall do this frequently in what follows and will not bother to explain this each time.

It is also worth pointing out that in some of the example below calculations involving vectors and/or linear transformation are carried out without reference to their components. One might have expected such examples to have been presented in Chapter 2. They are contained in the present chapter because they all involve either the determinant or trace of a linear transformation, and we chose to define these quantities in terms of components (even though they are basis independent).

Example 3.1: Suppose that **A** is a symmetric linear transformation. Show that its matrix of components $[A]$ in any basis is a symmetric matrix.

Solution: According to (3.13), the components of $\mathbf{A}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are defined by

$$A_{ji} = \mathbf{e}_j \cdot \mathbf{A}\mathbf{e}_i. \tag{i}$$

The property (NNN) of the transpose shows that $\mathbf{e}_j \cdot \mathbf{A}\mathbf{e}_i = \mathbf{A}^T\mathbf{e}_j \cdot \mathbf{e}_i$, which, on using the fact that $\mathbf{A}$ is symmetric further simplifies to $\mathbf{e}_j \cdot \mathbf{A}\mathbf{e}_i = \mathbf{A}\mathbf{e}_j \cdot \mathbf{e}_i$; and finally since the order of the vectors in a scalar product do not matter we have $\mathbf{e}_j \cdot \mathbf{A}\mathbf{e}_i = \mathbf{e}_i \cdot \mathbf{A}\mathbf{e}_j$. Thus

$$A_{ji} = \mathbf{e}_i \cdot \mathbf{A}\mathbf{e}_j\ . \tag{ii}$$

By (3.13), the right most term here is the A_{ij} component of $\mathbf{A}$, and so (ii) yields

$$A_{ji} = A_{ij}. \tag{iii}$$

Thus $[A] = [A]^T$ and so the matrix $[A]$ is symmetric.

Remark: Conversely, if it is known that the matrix of components $[A]$ of a linear transformation in some basis is is symmetric, then the linear transformation $\mathbf{A}$ is also symmetric.

Example 2.5: Choose any convenient basis and calculate the components of the projection linear transformation $\mathbf{\Pi}$ and the reflection linear transformation $\mathbf{R}$ in that basis.

Solution: Let $\mathbf{e}_3$ be a unit vector normal to the plane $\mathcal{P}$ and let $\mathbf{e}_1$ and $\mathbf{e}_2$ be any two unit vectors in $\mathcal{P}$ such that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms an orthonormal basis. From an example in the previous chapter we know that the projection transformation $\mathbf{\Pi}$ and the reflection transformation $\mathbf{R}$ in the plane $\mathcal{P}$ can be written as $\mathbf{\Pi} = \mathbf{I} - \mathbf{e}_3 \otimes \mathbf{e}_3$ and $\mathbf{R} = \mathbf{I} - 2(\mathbf{e}_3 \otimes \mathbf{e}_3)$ respectively. Since the components of $\mathbf{e}_3$ in the chosen basis are δ_{3i}, we find that

$$\Pi_{ij} = \delta_{ij} - (\mathbf{e}_3)_i(\mathbf{e}_3)_j = \delta_{ij} - \delta_{3i}\delta_{3j}, \qquad R_{ij} = \delta_{ij} - 2\delta_{3i}\delta_{3j}.$$

Example 3.2: Consider the scalar-valued function

$$f(\mathbf{A}, \mathbf{B}) \ = \ \text{trace}(\mathbf{A}\mathbf{B}^T) \tag{i}$$

and show that, for all linear transformations $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and for all scalars α, this function$f$ has the following properties:

i) $f(\mathbf{A}, \mathbf{B}) = f(\mathbf{B}, \mathbf{A})$,

ii) $f(\alpha\mathbf{A}, \mathbf{B}) = \alpha f(\mathbf{A}, \mathbf{B})$,

iii) $f(\mathbf{A} + \mathbf{C}, \mathbf{B}) = f(\mathbf{A}, \mathbf{B}) + f(\mathbf{C}, \mathbf{B})$ and

iv) $f(\mathbf{A}, \mathbf{A}) > 0$ provided $\mathbf{A} \neq \mathbf{0}$.

Solution: Let A_{ij} and B_{ij} be the components of $\mathbf{A}$ and $\mathbf{B}$ in an arbitrary basis. In terms of these components, $(\mathbf{A}\mathbf{B}^T)_{ij} = A_{ik}B^T_{kj} = A_{ik}B_{jk}$ and so

$$f(\mathbf{A}, \mathbf{B}) = A_{ik}B_{ik}\ . \tag{ii}$$

It is now trivial to verify that all of the above requirements hold.

Remark: It follows from this that the function f has all of the usual requirements of a scalar product. Therefore we may define the scalar-product of two linear transformations $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A}\cdot\mathbf{B}$, as

$$\mathbf{A}\cdot\mathbf{B} = \text{ trace}(\mathbf{A}\mathbf{B}^T) = A_{ij}B_{ij}. \tag{iii}$$

Note that, based on this scalar-product, we can define the magnitude of a linear transformation to be

$$|\mathbf{A}| = \sqrt{\mathbf{A}\cdot\mathbf{A}} = \sqrt{A_{ij}A_{ij}}. \tag{iv}$$

Example 3.3: For any two vectors $\mathbf{u}$ and $\mathbf{v}$, show that their cross-product $\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Solution: We are to show, for example, that $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v}) = 0$. In terms of their components we can write

$$\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v}) = u_i\ (\mathbf{u}\times\mathbf{v})_i = u_i\ (e_{ijk}u_jv_k) = e_{ijk}u_iu_jv_k\ . \tag{i}$$

Since $e_{ijk} = -e_{jik}$ and $u_iu_j = u_ju_i$, it follows that e_{ijk} is skew-symmetric in the subscripts ij and u_iu_j is symmetric in the subscripts ij. Thus it follows from Example 1.3 that $e_{ijk}u_iu_j = 0$ and so $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v}) = 0$. The orthogonality of $\mathbf{v}$ and $\mathbf{u}\times\mathbf{v}$ can be established similarly.

Example 3.4 Suppose that $\mathbf{a},\mathbf{b},\mathbf{c}$, are any three linearly independent vectors and that $\mathbf{F}$ be an arbitrary non-singular linear transformation. Show that

$$(\mathbf{Fa}\times\mathbf{Fb})\cdot\mathbf{Fc}\ =\ \det\mathbf{F}\ (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} \tag{i}$$

Solution: First consider the left-hand side of (i). On using (3.10), and (3.11), we can express this as

$$(\mathbf{Fa}\times\mathbf{Fb})\cdot\mathbf{Fc} = (\mathbf{Fa}\times\mathbf{Fb})_i\ (\mathbf{Fc})_i = e_{ijk}\ (\mathbf{Fa})_j\ (\mathbf{Fb})_k\ (\mathbf{Fc})_i, \tag{ii}$$

and consequently

$$(\mathbf{Fa}\times\mathbf{Fb})\cdot\mathbf{Fc} = e_{ijk}\ (F_{jp}\ a_p)\ (F_{kq}\ b_q)\ (F_{ir}\ c_r) = e_{ijk}\ F_{ir}F_{jp}F_{kq}a_p\ b_q\ c_r\ . \tag{iii}$$

Turning next to the right-hand side of (i), we note that

$$\det\mathbf{F}\ (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \det[F](\mathbf{a}\times\mathbf{b})_ic_i = \det[F]e_{ijk}a_jb_kc_i = \det[F]e_{rpq}a_pb_qc_r\ . \tag{iv}$$

Recalling the identity $e_{rpq}\det[F] = e_{ijk}F_{ir}F_{jp}F_{kq}$ in (1.48) for the determinant of a matrix and substituting this into (iv) gives

$$\det\mathbf{F}\ (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = e_{ijk}F_{ir}F_{jp}F_{kq}a_pb_qc_r. \tag{v}$$

Since the right-hand sides of (iii) and (v) are identical, it follows that the left-hand sides must also be equal, thus establishing the desired result.

Example 3.5 Suppose that $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ are three non-coplanar vectors in $\mathbb{E}_3$. Let V_0 be the volume of the tetrahedron defined by these three vectors. Next, suppose that $\mathbf{F}$ is a non-singular 2-tensor and let V denote the volume of the tetrahedron defined by the vectors $\mathbf{Fa}, \mathbf{Fb}$ and $\mathbf{Fc}$. Note that the second tetrahedron is the image of the first tetrahedron under the transformation $\mathbf{F}$. Derive a formula for V in terms of V_0 and $\mathbf{F}$.

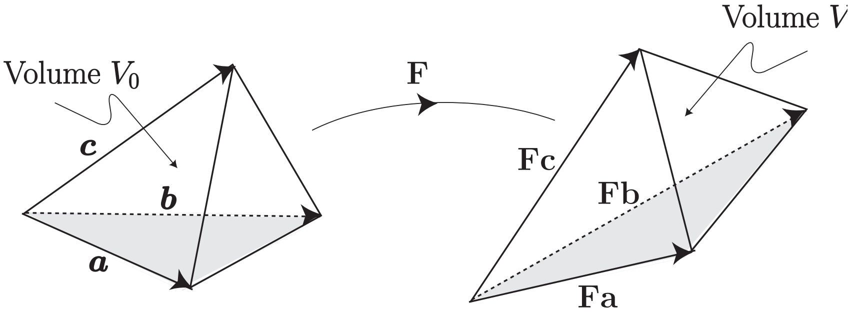


Figure 3.2: Tetrahedron of volume V_0 defined by three non-coplanar vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$; and its image under the linear transformation $\mathbf{F}$.

Solution: Recall from an example in the previous Chapter that the volume V_0 of the tetrahedron defined by any three non-coplanar vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is

$$V_0 = \frac{1}{6}(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.$$

The volume V of the tetrahedron defined by the three vectors $\mathbf{Fa}, \mathbf{Fb}, \mathbf{Fc}$ is likewise

$$V = \frac{1}{6}(\mathbf{Fa} \times \mathbf{Fb}) \cdot \mathbf{Fc}.$$

It follows from the result of the previous example that

$$V/V_0 = \det \mathbf{F}$$

which describes how volumes are mapped by the transformation $\mathbf{F}$.

Example 3.6: Suppose that $\mathbf{a}$ and $\mathbf{b}$ are two non-colinear vectors in $\mathbb{E}_3$. Let α_0 be the area of the parallelogram defined by these two vectors and let $\mathbf{n}_0$ be a unit vector that is normal to the plane of this parallelogram. Next, suppose that $\mathbf{F}$ is a non-singular 2-tensor and let α and $\mathbf{n}$ denote the area and unit normal to the parallelogram defined by the vectors $\mathbf{Fa}$ and $\mathbf{Fb}$. Derive formulas for α and $\mathbf{n}$ in terms of $\alpha_0, \mathbf{n}_0$ and $\mathbf{F}$.

Solution: By the properties of the vector-product we know that

$$\alpha_0 = |\mathbf{a} \times \mathbf{b}|, \qquad \mathbf{n}_0 = \frac{\mathbf{a} \times \mathbf{b}}{|\mathbf{a} \times \mathbf{b}|};$$

and similarly that

$$\alpha = |\mathbf{Fa} \times \mathbf{Fb}|, \qquad \mathbf{n} = \frac{\mathbf{Fa} \times \mathbf{Fb}}{|\mathbf{Fa} \times \mathbf{Fb}|}.$$

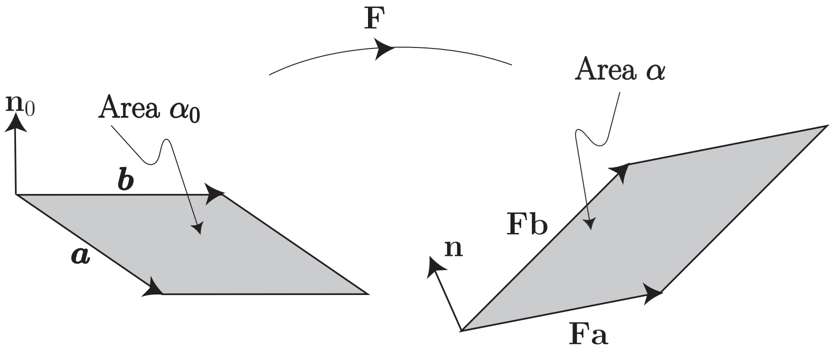


Figure 3.3: Parallelogram of area α_0 with unit normal $\mathbf{n}_0$ defined by two non-colinear vectors $\mathbf{a}$ and $\mathbf{b}$; and its image under the linear transformation $\mathbf{F}$.

Therefore

$$\alpha_0 \mathbf{n}_0 = \mathbf{a}\times\mathbf{b}, \qquad \text{and} \qquad \alpha\mathbf{n} = \mathbf{Fa}\times\mathbf{Fb}. \tag{i}$$

But

$$(\mathbf{Fa}\times\mathbf{Fb})_s = e_{sij}(\mathbf{Fa})_i(\mathbf{Fb})_j = e_{sij}F_{ip}a_pF_{jq}b_q. \tag{ii}$$

Also recall the identity $e_{pqr}\det[F] = e_{ijk}F_{ip}F_{jq}F_{kr}$ introduced in (1.48). Multiplying both sides of this identity by F^{-1}_{rs} leads to

$$e_{pqr}\det[F]F^{-1}_{rs} = e_{ijk}F_{ip}F_{jq}F_{kr}F^{-1}_{rs} = e_{ijk}F_{ip}F_{jq}\delta_{ks} = e_{ijs}F_{ip}F_{jq} = e_{sij}F_{ip}F_{jq} \tag{iii}$$

Substituting (iii) into (ii) gives

$$(\mathbf{Fa}\times\mathbf{Fb})_s = \det[F]e_{pqr}F^{-1}_{rs}a_pb_q = \det[F]e_{rpq}a_pb_qF^{-1}_{rs} = \det\mathbf{F}(\mathbf{a}\times\mathbf{b})_rF^{-T}_{sr} = \det\mathbf{F}\Big(\mathbf{F}^{-T}(\mathbf{a}\times\mathbf{b})\Big)_s$$

and so using (i),

$$\alpha\mathbf{n} = \alpha_0\det\mathbf{F}(\mathbf{F}^{-T}\mathbf{n}_0).$$

This describes how (vectorial) areas are mapped by the transformation $\mathbf{F}$. Taking the norm of this vector equation gives

$$\frac{\alpha}{\alpha_0} = |\det\mathbf{F}|\;|\mathbf{F}^{-T}\mathbf{n}_0|;$$

and substituting this result into the preceding equation gives

$$\mathbf{n} = \frac{\mathbf{F}^{-T}\mathbf{n}_0}{|\mathbf{F}^{-T}\mathbf{n}_0|}.$$

Example 3.5: Let $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ and $\{\mathbf{e}'_1,\mathbf{e}'_2,\mathbf{e}'_3\}$ be two bases related by nine scalars Q_{ij} through $\mathbf{e}'_i = Q_{ij}\mathbf{e}_j$. Let $\mathbf{Q}$ be the linear transformation whose components in the basis $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ are Q_{ij}. Show that

$$\mathbf{e}'_i = \mathbf{Q}^T\mathbf{e}_i;$$

thus $\mathbf{Q}^T$ is the transformation that carries the first basis into the second.

Solution: Since Q_{ij} are the components of the linear transformation $\mathbf{Q}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, it follows from the definition of components that

$$\mathbf{Q}\mathbf{e}_j = Q_{ij}\mathbf{e}_i.$$

Since $[Q]$ is an orthogonal matrix one readily sees that $\mathbf{Q}$ is an orthogonal transformation. Operating on both sides of the preceding equation by $\mathbf{Q}^T$ and using the orthogonality of $\mathbf{Q}$ leads to

$$\mathbf{e}_j = Q_{ij}\mathbf{Q}^T\mathbf{e}_i.$$

Multiplying both sides of this by Q_{kj} and noting by the orthogonality of $\mathbf{Q}$ that $Q_{kj}Q_{ij} = \delta_{ki}$, we are now led to

$$Q_{kj}\mathbf{e}_j = \mathbf{Q}^T\mathbf{e}_k$$

or equivalently

$$\mathbf{Q}^T\mathbf{e}_i = Q_{ij}\mathbf{e}_j.$$

This, together with the given fact that $\mathbf{e}'_i = Q_{ij}\mathbf{e}_j$, yields the desired result.

Example 3.6: Determine the relationship between the components v_i and v'_i of a vector $\mathbf{v}$ in two bases.

Solution: The components v_i of $\mathbf{v}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are defined by

$$v_i = \mathbf{v} \cdot \mathbf{e}_i,$$

and its components v'_i in the second basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ are defined by

$$v'_i = \mathbf{v} \cdot \mathbf{e}'_i.$$

It follows from this and (NNN) that

$$v'_i = \mathbf{v} \cdot \mathbf{e}'_i = \mathbf{v} \cdot (Q_{ij}\mathbf{e}_j) = Q_{ij}\mathbf{v} \cdot \mathbf{e}_j = Q_{ij}v_j.$$

Thus, the components of the vector $\mathbf{v}$ in the two bases are related by

$$v'_i = Q_{ij}v_j.$$

Example 3.7: Determine the relationship between the components A_{ij} and A'_{ij} of a linear transformation $\mathbf{A}$ in two bases.

Solution: The components A_{ij} of the linear transformation $\mathbf{A}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are defined by

$$A_{ij} = \mathbf{e}_i \cdot (\mathbf{A}\mathbf{e}_j), \tag{i}$$

and its components A'_{ij} in a second basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ are defined by

$$A'_{ij} = \mathbf{e}'_i \cdot (\mathbf{A}\mathbf{e}'_j). \tag{ii}$$

By first making use of (NNN), and then (i), we can write (ii) as

$$A'_{ij} = \mathbf{e}'_i \cdot (\mathbf{A}\mathbf{e}'_j) = Q_{ip}\mathbf{e}_p \cdot (\mathbf{A}Q_{jq}\mathbf{e}_q) = Q_{ip}Q_{jq}\mathbf{e}_p \cdot (\mathbf{A}\mathbf{e}_q) = Q_{ip}Q_{jq}A_{pq}. \tag{iii}$$

Thus, the components of the linear transformation $\mathbf{A}$ in the two bases are related by

$$A'_{ij} = Q_{ip}Q_{jq}A_{pq}. \tag{iv}$$

Example 3.8: Suppose that the basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ is obtained by rotating the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ through an angle θ about the unit vector $\mathbf{e}_3$; see Figure 3.4. Write out the transformation rule for 2-tensors explicitly in this case.

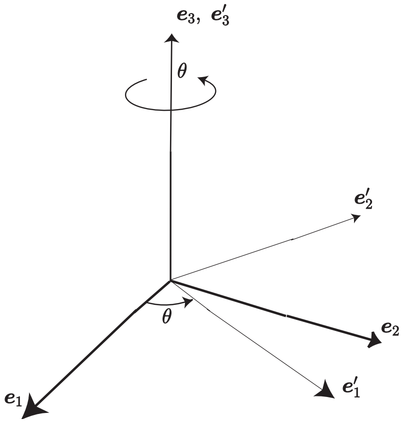


Figure 3.4: A basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ obtained by rotating the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ through an angle θ about the unit vector $\mathbf{e}_3$.

Solution: In view of the given relationship between the two bases it follows that

$$\left.\begin{aligned} \mathbf{e}'_1 &= \cos\theta\ \mathbf{e}_1 + \sin\theta\ \mathbf{e}_2, \\ \mathbf{e}'_2 &= -\sin\theta\ \mathbf{e}_1 + \cos\theta\ \mathbf{e}_2, \\ \mathbf{e}'_3 &= \mathbf{e}_3. \end{aligned}\right\}$$

The matrix $[Q]$ which relates the two bases is defined by $Q_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j$, and so it follows that

$$[Q] = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Substituting this $[Q]$ into $[A'] = [Q][A][Q]^T$ and multiplying out the matrices leads to the 9 equations

$$\begin{aligned}
A'_{11} &= \frac{A_{11}+A_{22}}{2} + \frac{A_{11}-A_{22}}{2}\cos 2\theta + \frac{A_{12}+A_{21}}{2}\sin 2\theta,\\
A'_{12} &= \frac{A_{12}-A_{21}}{2} + \frac{A_{12}-A_{21}}{2}\cos 2\theta - \frac{A_{11}-A_{22}}{2}\sin 2\theta,\\
A'_{21} &= -\frac{A_{12}-A_{21}}{2} - \frac{A_{12}-A_{21}}{2}\cos 2\theta - \frac{A_{11}-A_{22}}{2}\sin 2\theta,\\
A'_{22} &= \frac{A_{11}+A_{22}}{2} - \frac{A_{11}-A_{22}}{2}\cos 2\theta - \frac{A_{12}+A_{21}}{2}\sin 2\theta,\\
A'_{13} &= A_{13}\cos\theta + A_{23}\sin\theta, \qquad A'_{31} = A_{31}\cos\theta + A_{32}\sin\theta,\\
A'_{23} &= A_{23}\cos\theta - A_{13}\sin\theta, \qquad A'_{32} = A_{32}\cos\theta - A_{31}\sin\theta,\\
A'_{33} &= A_{33}\ .
\end{aligned}$$

In the special case when $[A]$ is symmetric, and in addition $A_{13} = A_{23} = 0$, these nine equations simplify to

$$\left.\begin{aligned}
A'_{11} &= \frac{A_{11}+A_{22}}{2} + \frac{A_{11}-A_{22}}{2}\cos 2\theta + A_{12}\sin 2\theta,\\
A'_{22} &= \frac{A_{11}+A_{22}}{2} - \frac{A_{11}-A_{22}}{2}\cos 2\theta - A_{12}\sin 2\theta,\\
A'_{12} &= -\frac{A_{11}-A_{22}}{2}\sin 2\theta,
\end{aligned}\right\}$$

together with $A'_{13} = A'_{23} = 0$ and $A'_{33} = A_{33}$. These are the well-known equations underlying the *Mohr's circle* for transforming 2-tensors in two-dimensions.

Example 3.9:

a. Let $\mathbf{a}, \mathbf{b}$ and $\mathcal{T}$ be entities whose components in some arbitrary basis are a_i, b_i and $\mathcal{T}_{ij}$. The components of $\mathcal{T}$ in any basis are defined in terms of the components of $\mathbf{a}$ and $\mathbf{b}$ in that basis by

$$\mathcal{T}_{ijk} = a_i b_j b_k. \tag{i}$$

If $\mathbf{a}$ and $\mathbf{b}$ are vectors, show that $\mathcal{T}$ is a 3-tensor.

b. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are 2-tensors and that their components in some basis are related by

$$A_{ij} = \mathbb{C}_{ijk\ell} B_{k\ell}. \tag{ii}$$

Show that the $\mathbb{C}_{ijk\ell}$'s are the components of a 4-tensor.

Solution:

a. Let a_i, a'_i and b_i, b'_i be the components of $\mathbf{a}$ and $\mathbf{b}$ in two arbitrary bases. We are told that the components of the entity $\mathcal{T}$ in these two bases are defined by

$$\mathcal{T}_{ijk} = a_i b_j b_k, \qquad \mathcal{T}'_{ijk} = a'_i b'_j b'_k. \tag{iii}$$

Since **a** and **b** are known to be vectors, their components transform according to the 1-tensor transformation rule

$$a'_i = Q_{ij} a_j, \qquad b'_i = Q_{ij} b_j. \tag{iv}$$

Combining equations (iii) and (iv) gives

$$\mathcal{T}'_{ijk} = a'_i b'_j b'_k = Q_{ip} a_p Q_{jq} b_q Q_{kr} b_r = Q_{ip} Q_{jq} Q_{kr} a_p b_q b_r = Q_{ip} Q_{jq} Q_{kr} \mathcal{T}_{pqr}. \tag{v}$$

Therefore the components of $\mathcal{T}$ in two bases transform according to $\mathcal{T}'_{ijk} = Q_{ip} Q_{jq} Q_{kr} \mathcal{T}_{pqr}$. Therefore $\mathcal{T}$ is a 3-tensor.

b. Let $A_{ij}, B_{ij}, \mathbb{C}_{ijk\ell}$ and $A'_{ij}, B'_{ij}, \mathbb{C}'_{ijk\ell}$ be the components of $\mathbf{A}, \mathbf{B}, \mathbb{C}$ in two arbitrary bases:

$$A_{ij} = \mathbb{C}_{ijk\ell} B_{k\ell}, \qquad A'_{ij} = \mathbb{C}'_{ijk\ell} B'_{k\ell}. \tag{vi}$$

We are told that **A** and **B** are 2-tensors, whence

$$A'_{ij} = Q_{ip} Q_{jq} A_{pq}, \qquad B'_{ij} = Q_{ip} Q_{jq} B_{pq}, \tag{vii}$$

and we must show that $\mathbb{C}_{ijk\ell}$ is a 4-tensor, i.e that $\mathbb{C}'_{ijk\ell} = Q_{ip} Q_{jq} Q_{kr} Q_{\ell s} \mathbb{C}_{pqrs}$. Substituting (vii) into (vi)$_2$ gives

$$Q_{ip} Q_{jq} A_{pq} = \mathbb{C}'_{ijk\ell} Q_{kp} Q_{\ell q} B_{pq}, \tag{viii}$$

Multiplying both sides by $Q_{im} Q_{jn}$ and using the orthogonality of $[Q]$, i.e. the fact that $Q_{ip} Q_{im} = \delta_{pm}$, leads to

$$\delta_{pm} \delta_{qn} A_{pq} = \mathbb{C}'_{ijk\ell} Q_{im} Q_{jn} Q_{kp} Q_{\ell q} B_{pq}, \tag{ix}$$

which by the substitution rule tells us that

$$A_{mn} = \mathbb{C}'_{ijk\ell} Q_{im} Q_{jn} Q_{kp} Q_{\ell q} B_{pq}, \tag{x}$$

or on using (vi)$_1$ in this that

$$\mathbb{C}_{mnpq} B_{pq} = \mathbb{C}'_{ijk\ell} Q_{im} Q_{jn} Q_{kp} Q_{\ell q} B_{pq}. \tag{xi}$$

Since this holds for all matrices $[B]$ we must have

$$\mathbb{C}_{mnpq} = \mathbb{C}'_{ijk\ell} Q_{im} Q_{jn} Q_{kp} Q_{\ell q}. \tag{xii}$$

Finally multiplying both sides by $Q_{am} Q_{bn} Q_{cp} Q_{dq}$, using the orthogonality of $[Q]$ and the substitution rule yields the desired result

$$Q_{am} Q_{bn} Q_{cp} Q_{dq} \mathbb{C}_{mnpq} = \mathbb{C}'_{abcd}. \tag{xiii}$$

Example 3.10: Verify that the alternator e_{ijk} has the property that

$$e_{ijk} = Q_{ip}\ Q_{jq} Q_{kr}\ e_{pqr} \quad \text{for all } \textit{proper} \text{ orthogonal matrices } [Q], \tag{i}$$

but that more generally

$$e_{ijk} \neq Q_{ip}\ Q_{jq} Q_{kr}\ e_{pqr} \quad \text{for } \textit{all} \text{ orthogonal matrices } [Q]. \tag{ii}$$

Note from this that the alternator is not an isotropic 3-tensor.

Example 3.11: If $\mathbb{C}_{ijk\ell}$ is an isotropic 4-tensor, show that necessarily $\mathbb{C}_{iik\ell} = \alpha\delta_{k\ell}$ for some arbitrary scalar α.

Solution: Since $\mathbb{C}_{ijkl}$ is an isotropic 4-tensor, by definition,

$$\mathbb{C}_{ijkl} = Q_{ip}\, Q_{jq}\, Q_{kr}\, Q_{ls}\, \mathbb{C}_{pqrs}$$

for all orthogonal matrices $[Q]$. On setting $i = j$ in this; then using the orthogonality of $[Q]$; and finally using the substitution rule, we are led to

$$\mathbb{C}_{iikl} = Q_{ip}\, Q_{iq}\, Q_{kr}\, Q_{ls}\, \mathbb{C}_{pqrs} = \delta_{pq}\, Q_{kr}\, Q_{ls}\, \mathbb{C}_{pqrs} = Q_{kr}\, Q_{ls}\, \mathbb{C}_{pprs}\ .$$

Thus $\mathbb{C}_{iik\ell}$ obeys

$$\mathbb{C}_{iikl} = Q_{kr}\, Q_{ls}\, \mathbb{C}_{pprs} \qquad \text{for all orthogonal matrices } [Q],$$

and therefore it is an isotropic 2-tensor. The desired result now follows since the most general isotropic 2-tensor is a scalar multiple of the identity.

Example 3.12: Show that the most general isotropic vector is the null vector $\mathbf{o}$.

Solution: In order to show this we must determine the most general vector $\mathbf{u}$ which is such that

$$u_i = Q_{ij} u_j \quad \text{for } \textit{all} \text{ orthogonal matrices } [Q]. \tag{i}$$

Since (i) is to hold for all orthogonal matrices $[Q]$, it must necessarily hold for the special choice $[Q] = -[I]$. Then $Q_{ij} = -\delta_{ij}$, and so (i) reduces to

$$u_i = -\delta_{ij} u_j = -u_i; \tag{ii}$$

thus $u_i = 0$ and so $\mathbf{u} = \mathbf{o}$.

Conversely, $\mathbf{u} = \mathbf{o}$ obviously satisfies (i) for all orthogonal matrices $[Q]$. Thus $\mathbf{u} = \mathbf{o}$ is the most general isotropic vector.

Example 3.13: Show that the most general isotropic symmetric tensor is a scalar multiple of the identity.

Solution: We must find the most general symmetric 2-tensor $\mathbf{A}$ whose components in every basis are the same; i.e.,

$$[A] = [Q][A][Q]^T \text{ for all orthogonal matrices } [Q]. \tag{i}$$

First, since $\mathbf{A}$ is symmetric, we know that there is some basis in which $[A]$ is diagonal. Since $\mathbf{A}$ is also isotropic, it follows that $[A]$ must therefore be diagonal in every basis. Thus $[A]$ has the form

$$[A] = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \tag{ii}$$

in any basis. Thus (i) takes the form

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} = [Q] \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} [Q^T] \qquad \text{(iii)}$$

for all orthogonal matrices $[Q]$. Thus (iii) must necessarily hold for the special choice

$$[Q] = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \text{(iv)}$$

in which case (iii) reduces to

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} = \begin{pmatrix} \lambda_2 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \qquad \text{(v)}$$

Therefore, $\lambda_1 = \lambda_2$.

A permutation of this special choice of $[Q]$ similarly shows that $\lambda_2 = \lambda_3$. Thus $\lambda_1 = \lambda_2 = \lambda_3 =$ say α. Therefore $[A]$ necessarily must have the form $[A] = \alpha[I]$.

Conversely, by direct substitution, $[A] = \alpha[I]$ is readily shown to obey (i) for any orthogonal matrix $[Q]$. This establishes the result.

Example 3.14: If $\mathbf{W}$ is a skew-symmetric tensor, show that there is a vector $\mathbf{w}$ such that $\mathbf{Wx} = \mathbf{w} \times \mathbf{x}$ for all $\mathbf{x} \in \mathbb{E}$.

Solution: Let W_{ij} be the components of $\mathbf{W}$ in some basis and let $\mathbf{w}$ be the vector whose components in this basis are defined by

$$w_i = -\frac{1}{2}\, e_{ijk} W_{jk}. \qquad \text{(i)}$$

Then, we merely have to show that $\mathbf{w}$ has the desired property stated above.

Multiplying both sides of the preceding equation by e_{ipq} and then using the identity $e_{ijk}e_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}$, and finally using the substitution rule gives

$$e_{ipq} w_i = -\frac{1}{2}\left(\delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}\right) W_{jk} = -\frac{1}{2}\left(W_{pq} - W_{qp}\right)$$

Since $\mathbf{W}$ is skew-symmetric we have $W_{ij} = -W_{ji}$ and thus conclude that

$$W_{ij} = -e_{ijk} w_k.$$

Now for any vector $\mathbf{x}$,

$$W_{ij} x_j = -e_{ijk} w_k x_j = e_{ikj} w_k x_j = (\mathbf{w} \times \mathbf{x})_i.$$

Thus the vector $\mathbf{w}$ defined by (i) has the desired property $\mathbf{Wx} = \mathbf{w} \times \mathbf{x}$.

Example 3.15: Verify that the 4-tensor

$$\mathbb{C}_{ijk\ell} = \alpha\delta_{ij}\delta_{k\ell} + \beta\delta_{ik}\delta_{j\ell} + \gamma\delta_{i\ell}\delta_{jk}, \tag{i}$$

where α, β, γ are scalars, is isotropic. If this isotropic 4-tensor is to possess the symmetry $\mathbb{C}_{ijk\ell} = \mathbb{C}_{jik\ell}$, show that one must have $\beta = \gamma$.

Solution: In order to verify that $\mathbb{C}_{ijk\ell}$ are the components of an isotropic 4-tensor we have to show that $\mathbb{C}_{ijk\ell} = Q_{ip}\, Q_{jq}\, Q_{kr}\, Q_{\ell s}\, \mathbb{C}_{pqrs}$ for all orthogonal matrices $[Q]$. The right-hand side of this can be simplified by using the given form of $\mathbb{C}_{ijk\ell}$; the substitution rule; and the orthogonality of $[Q]$ as follows:

$$\begin{aligned}
&Q_{ip}\, Q_{jq}\, Q_{kr}\, Q_{\ell s}\, \mathbb{C}_{pqrs}\\
&\quad= Q_{ip}\, Q_{jq}\, Q_{kr}\, Q_{\ell s}(\alpha\, \delta_{pq}\, \delta_{rs} + \beta\, \delta_{pr}\, \delta_{qs} + \gamma\, \delta_{ps}\, \delta_{qr})\\
&\quad= \alpha\, Q_{iq}\, Q_{jq}\, Q_{ks}\, Q_{\ell s} + \beta\, Q_{ir}\, Q_{js}\, Q_{kr}\, Q_{\ell s} + \gamma\, Q_{is}\, Q_{jr}\, Q_{kr}\, Q_{\ell s}\\
&\quad= \alpha\,(Q_{iq}Q_{jq})\,(Q_{ks}Q_{\ell s}) + \beta\,(Q_{ir}Q_{kr})\,(Q_{js}Q_{\ell s}) + \gamma\,(Q_{is}Q_{\ell s})\,(Q_{jr}Q_{kr})\\
&\quad= \alpha\, \delta_{ij}\, \delta_{k\ell} + \beta\, \delta_{ik}\, \delta_{j\ell} + \gamma\, \delta_{i\ell}\, \delta_{jk}\\
&\quad= \mathbb{C}_{ijk\ell}.
\end{aligned} \tag{ii}$$

This establishes the desired result.

Turning to the second question, enforcing the requirement $\mathbb{C}_{ijk\ell} = \mathbb{C}_{jik\ell}$ on (i) leads, after some simplification, to

$$(\beta - \gamma)\,(\delta_{ik}\,\delta_{j\ell} - \delta_{jk}\,\delta_{i\ell}) = 0\ . \tag{iii}$$

Since this must hold for all values of the free indices i, j, k, ℓ, it must necessarily hold for the special choice $i = 1,\ j = 2,\ k = 1,\ \ell = 2$. Therefore $(\beta - \gamma)(\delta_{11}\,\delta_{22} - \delta_{21}\,\delta_{12}) = 0$ and so

$$\beta = \gamma. \tag{iv}$$

Remark: We have shown that $\beta = \gamma$ is *necessary* if $\mathbb{C}$ given in (i) is to have the symmetry $\mathbb{C}_{ijk\ell} = \mathbb{C}_{jik\ell}$. One can readily verify that it is *sufficient* as well. It is useful for later use to record here, that the most general isotropic 4-tensor $\mathbb{C}$ with the symmetry property $\mathbb{C}_{ijk\ell} = \mathbb{C}_{jik\ell}$ is

$$\mathbb{C}_{ijk\ell} = \alpha\delta_{ij}\delta_{k\ell} + \beta\,(\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}) \tag{v}$$

where α and β are scalars.

Remark: Observe that $\mathbb{C}_{ijk\ell}$ given by (v) automatically has the symmetry $\mathbb{C}_{ijk\ell} = \mathbb{C}_{k\ell ij}$.

Example 3.16: If $\mathbf{A}$ is a tensor such that

$$\mathbf{Ax} \cdot \mathbf{x} = 0 \qquad \text{for all } \mathbf{x} \tag{i}$$

show that $\mathbf{A}$ is necessarily skew-symmetric.

Solution: By definition of the transpose and the properties of the scalar product, $\mathbf{Ax}\cdot\mathbf{x} = \mathbf{x}\cdot\mathbf{A}^T\mathbf{x} = \mathbf{A}^T\mathbf{x}\cdot\mathbf{x}$. Therefore $\mathbf{A}$ has the properties that

$$\mathbf{Ax}\cdot\mathbf{x} = 0, \quad \text{and} \quad \mathbf{A}^T\mathbf{x}\cdot\mathbf{x} = 0 \qquad \text{for all vectors } \mathbf{x}.$$

Adding these two equations gives

$$\mathbf{S}\mathbf{x}\cdot\mathbf{x} = 0 \qquad \text{where} \quad \mathbf{S} = \Big(\mathbf{A} + \mathbf{A}^T\Big) .$$

Observe that $\mathbf{S}$ is symmetric. Therefore in terms of components in a principal basis of $\mathbf{S}$,

$$\mathbf{S}\mathbf{x}\cdot\mathbf{x} = \sigma_1 x_1^2 + \sigma_2 x_2^2 + \sigma_3 x_3^2 = 0$$

where the σ_k's are the eigenvalues of $\mathbf{S}$. Since this must hold for all real numbers x_k, it follows that every eigenvalue must vanish: $\sigma_1 = \sigma_2 = \sigma_3 = 0$. Therefore $\mathbf{S} = \mathbf{O}$ whence

$$\mathbf{A} = -\mathbf{A}^T.$$

Remark: An important consequence of this is that if $\mathbf{A}$ is a tensor with the property that $\mathbf{A}\mathbf{x}\cdot\mathbf{x} = 0$ for all $\mathbf{x}$, it does *not* follow that $\mathbf{A} = \mathbf{0}$ necessarily.

Example 2.18: For any orthogonal linear transformation $\mathbf{Q}$, show that $\det\mathbf{Q} = \pm 1$.

Solution: Recall that for any two linear transformations $\mathbf{A}$ and $\mathbf{B}$ we have $\det(\mathbf{AB}) = \det\mathbf{A}\det\mathbf{B}$ and $\det\mathbf{B} = \det\mathbf{B}^T$. Since $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ it now follows that $1 = \det\mathbf{I} = \det(\mathbf{Q}\mathbf{Q}^T) = \det\mathbf{Q}\det\mathbf{Q}^T = (\det\mathbf{Q})^2$. The desired result now follows.

Example 2.20: If $\mathbf{Q}$ is a *proper* orthogonal linear transformation on the vector space $\mathbb{E}_3$, show that there exists a vector $\mathbf{v}$ such that $\mathbf{Q}\mathbf{v} = \mathbf{v}$. This vector is known as the axis of $\mathbf{Q}$.

Solution: To show that there is a vector $\mathbf{v}$ such that $\mathbf{Q}\mathbf{v} = \mathbf{v}$, it is sufficient to show that $\mathbf{Q}$ has an eigenvalue $+1$, i.e. that $(\mathbf{Q} - \mathbf{I})\mathbf{v}) = \mathbf{o}$ or equivalently that $\det(\mathbf{Q} - \mathbf{I}) = 0$.

Since $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ we have $\mathbf{Q}(\mathbf{Q}^T - \mathbf{I}) = \mathbf{I} - \mathbf{Q}$. On taking the determinant of both sides and using the fact that $\det(\mathbf{AB}) = \det\mathbf{A}\det\mathbf{B}$ we get

$$\det\mathbf{Q}\ \det(\mathbf{Q}^T - \mathbf{I}) = \det\left(\mathbf{I} - \mathbf{Q}\right)\ . \tag{i}$$

Recall that $\det\mathbf{Q} = +1$ for a proper orthogonal linear transformation, and that $\det\mathbf{A} = \mathbf{A}^T$ and $\det(-\mathbf{A}) = (-1)^3\det(\mathbf{A})$ for a 3-dimensional vector space. Therefore this leads to

$$\det(\mathbf{Q} - \mathbf{I}) = -\det(\mathbf{Q} - \mathbf{I}), \tag{ii}$$

and the desired result now follows.

Example 2.22: For any linear transformation $\mathbf{A}$, show that $\det(\mathbf{A} - \mu\mathbf{I}) = \det(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \mu\mathbf{I})$ for all orthogonal linear transformations $\mathbf{Q}$ and all scalars μ.

Solution: This follows readily since

$$\det(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \mu\mathbf{I}) = \det(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \mu\mathbf{Q}^T\mathbf{Q}) = \det\Big(\mathbf{Q}^T(\mathbf{A} - \mu\mathbf{I})\mathbf{Q}\Big) = \det\mathbf{Q}^T\det(\mathbf{A} - \mu\mathbf{I})\det\mathbf{Q} = \det(\mathbf{A} - \mu\mathbf{I}).$$

Remark: Observe from this result that the eigenvalues of $\mathbf{Q}^T\mathbf{A}\mathbf{Q}$ coincide with those of $\mathbf{Q}$, so that in particular the same is true of their product and their sum: $\det(\mathbf{Q}^T\mathbf{A}\mathbf{Q}) = \det\mathbf{A}$ and $\mathrm{tr}(\mathbf{Q}^T\mathbf{A}\mathbf{Q}) = \mathrm{tr}\ \mathbf{A}$.

Example 2.26: Define a scalar-valued function $\phi(\mathbf{A};\ \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ for all linear transformations $\mathbf{A}$ and all (not necessarily) orthonormal bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ by

$$\phi(\mathbf{A};\ \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \frac{\mathbf{A}\mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3) + \mathbf{e}_1 \cdot (\mathbf{A}\mathbf{e}_2 \times \mathbf{e}_3) + \mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{A}\mathbf{e}_3)}{\mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3)}.$$

Show that $\phi(\mathbf{A},\ \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is in fact independent of the choice of basis, i.e., show that

$$\phi(\mathbf{A}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = I_1(\mathbf{A}, \mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)$$

for any two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$. Thus, we can simply write $\phi(\mathbf{A})$ instead of $\phi(\mathbf{A}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$; $\phi(\mathbf{A})$ is called a *scalar invariant* of $\mathbf{A}$.

Pick any orthonormal basis and express $\phi(\mathbf{A})$ in terms of the components of $\mathbf{A}$ in that basis; and hence show that $\phi(\mathbf{A}) = \text{trace}\ \mathbf{A}$.

Example 3.7: Let $\mathbf{F}(t)$ be a one-parameter familty of non-singular 2-tensors that depends smoothly on the parameter t. Calculate

$$\frac{\mathrm{d}}{\mathrm{dt}} \det\mathbf{F}(t).$$

Solution: From the result of Example 3.NNN we have

$$\big(\mathbf{F}(t)\mathbf{a} \times \mathbf{F}(t)\mathbf{b}\big) \cdot \mathbf{F}(t)\mathbf{c} \ = \ \det\mathbf{F}(t)\ (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$$

Differentiating this with respect to t gives

$$\big(\dot{\mathbf{F}}(t)\mathbf{a} \times \mathbf{F}(t)\mathbf{b}\big) \cdot \mathbf{F}(t)\mathbf{c} \ + \big(\mathbf{F}(t)\mathbf{a} \times \dot{\mathbf{F}}(t)\mathbf{b}\big) \cdot \mathbf{F}(t)\mathbf{c} \ + \big(\mathbf{F}(t)\mathbf{a} \times \mathbf{F}(t)\mathbf{b}\big) \cdot \dot{\mathbf{F}}(t)\mathbf{c} \ = \ \frac{\mathrm{d}}{\mathrm{dt}} \det\mathbf{F}(t)\ (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$$

where we have set $\dot{\mathbf{F}}(t) = d\mathbf{F}/dt$. We can write this as

$$\big(\dot{\mathbf{F}}\mathbf{F}^{-1}\mathbf{F}\mathbf{a} \times \mathbf{F}\mathbf{b}\big) \cdot \mathbf{F}\mathbf{c} + \big(\mathbf{F}\mathbf{a} \times \dot{\mathbf{F}}\mathbf{F}^{-1}\mathbf{F}\mathbf{b}\big) \cdot \mathbf{F}\mathbf{c} + \big(\mathbf{F}\mathbf{a} \times \mathbf{F}\mathbf{b}\big) \cdot \dot{\mathbf{F}}\mathbf{F}^{-1}\mathbf{F}\mathbf{c} \ = \left(\frac{\mathrm{d}}{\mathrm{dt}} \det\mathbf{F}\right)\ (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.$$

In view of the result of Example 3.NNN, this can be written as

$$\text{trace}\Big(\dot{\mathbf{F}}\mathbf{F}^{-1}\Big)\big(\mathbf{F}\mathbf{a} \times \mathbf{F}\mathbf{b}\big) \cdot \mathbf{F}\mathbf{c} \ = \left(\frac{\mathrm{d}}{\mathrm{dt}} \det\mathbf{F}\right)\ (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$$

and now using the result of Example 3.NNN once more we get

$$\text{trace}\Big(\dot{\mathbf{F}}\mathbf{F}^{-1}\Big) \det\mathbf{F}\ (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} \ = \left(\frac{\mathrm{d}}{\mathrm{dt}} \det\mathbf{F}\right)\ (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.$$

or

$$\frac{\mathrm{d}}{\mathrm{dt}} \det\mathbf{F} \ = \text{trace}\Big(\dot{\mathbf{F}}\mathbf{F}^{-1}\Big) \det\mathbf{F}.$$

Example 2.30: For any integer $N > 0$, show that the polynomial

$$P_N(\mathbf{A}) = c_0\mathbf{I} + c_1\mathbf{A} + c_2\mathbf{A}^2 + \ldots c_k\mathbf{A}^k + \ldots + c_N\mathbf{A}^N$$

can be written as a quadratic polynomial of $\mathbf{A}$.

Solution: This follows readily from the Cayley-Hamilton Theorem (3.41) as follows: suppose that $\mathbf{A}$ is non-singular so that $I_3(\mathbf{A}) = \det \mathbf{A} \neq 0$. Then (3.41) shows that $\mathbf{A}^3$ can be written as a linear combination of $\mathbf{I}$, $\mathbf{A}$ and $\mathbf{A}^2$. Next, multiplying this by $\mathbf{A}$ tells us that $\mathbf{A}^4$ can be written as a linear combination of $\mathbf{A}$, $\mathbf{A}^2$ and $\mathbf{A}^3$, and therefore, on using the result of the previous step, as linear combination of $\mathbf{I}$, $\mathbf{A}$ and $\mathbf{A}^2$. This process can be continued an arbitrary number of times to see that for any integer k, $\mathbf{A}^k$ can be expressed as a linear combination of $\mathbf{I}$, $\mathbf{A}$ and $\mathbf{A}^2$. The result thus follows.

Example 2.31: For any linear transformation $\mathbf{A}$ show that

$$\det(\mathbf{A} - \alpha\mathbf{I}) = -\alpha^3 + I_1(\mathbf{A})\alpha^2 - I_2(\mathbf{A})\alpha + I_3(\mathbf{A})$$

for *all* real numbers α where $I_1(\mathbf{A}), I_2(\mathbf{A})$ and $I_3(\mathbf{A})$ are the principal scalar invariants of $\mathbf{A}$:

$$I_1(\mathbf{A}) = \text{trace } \mathbf{A}, \qquad I_2(\mathbf{A}) = 1/2[(\text{trace } \mathbf{A})^2 - \text{trace}(\mathbf{A}^2)], \qquad I_3(\mathbf{A}) = \det \ \mathbf{A}.$$

Example 2.32: Calculate the principal scalar invariants I_1, I_2 and I_3 of the linear transformation $\mathbf{a} \otimes \mathbf{b}$.

References

1. H. Jeffreys, *Cartesian Tensors*, Cambridge, 1931.

2. J.K. Knowles, *Linear Vector Spaces and Cartesian Tensors*, Oxford University Press, New York, 1997.

3. L.A. Segel, *Mathematics Applied to Continuum Mechanics*, Dover, New York, 1987.

Chapter 4

Characterizing Symmetry: Groups of Linear Transformations.

Linear transformations are mappings of vector spaces into vector spaces. When an object is mapped using a linear transformation, certain transformations preserve its symmetry while others don't. One way in which to characterize the symmetry of an object is to consider the collection of all linear transformations that preserve its symmetry. The set of such transformations depends on the object: for example the set of linear transformations that preserve the symmetry of a cube is different to the set of linear transformations that preserve the symmetry of a tetrahedron. In this chapter we touch briefly on the question of characterizing symmetry by linear transformations.

Intuitively a "uniform all-around expansion", i.e. a linear transformation of the form $\alpha\mathbf{I}$ that rescales the object by changing its size but not its shape, does not affect symmetry. We are interested in other linear transformations that also preserve symmetry, principally rotations and reflections. In this Chapter we shall consider those linear transformations that *map the object back into itself.* The collection of such transformations have certain important and useful properties.

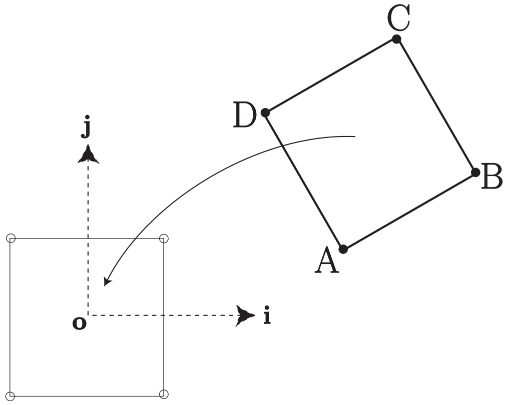


Figure 4.1: Mapping a square into itself.

4.1 An example in two-dimensions.

We begin with an illustrative example. Consider a square, ABCD, which lies in a plane normal to the unit vector **k**, whose center is at the origin **o** and whose sides are parallel to the orthonormal vectors $\{\mathbf{i}, \mathbf{j}\}$. Consider mappings that carry the square into itself. The vertex A can be placed in one of 4 positions; see Figure 4.1. Once the location of A has been determined, the vertex B can be placed in one of 2 positions (allowing for reflections or in just one position if only rotations are permitted). And once the locations of A and B have been fixed, there is no further flexibility and the locations of the remaining vertices are fixed. Thus there are a total of $4 \times 2 = 8$ symmetry preserving transformations of the square, 4 of which are rotations and 4 of which are reflections.

Consider the 4 rotations. In order to determine them, we (a) identify the axes of rotational symmetry and then (b) determine the number of distinct rotations about each such axis. In the present case there is just 1 axis to consider, viz. **k**, and we note that $0^o, 90^o, 180^o$ and 270^o rotations about this axis map the square back into itself. Thus the following 4 distinct rotations are symmetry transformations: $\mathbf{I}, \mathbf{R}_{\mathbf{k}}^{\pi/2},\ \mathbf{R}_{\mathbf{k}}^{\pi},\ \mathbf{R}_{\mathbf{k}}^{3\pi/2}$ where we are using the notation introduced previously, i.e. $\mathbf{R}_{\mathbf{n}}^{\phi}$ is a right-handed rotation through an angle ϕ about the axis **n**.

Let $\mathcal{G}_{\text{square}}$ denote the set consisting of these 4 symmetry preserving rotations:

$$\mathcal{G}_{\text{square}} = \{\mathbf{I},\ \mathbf{R}_{\mathbf{k}}^{\pi/2},\ \mathbf{R}_{\mathbf{k}}^{\pi},\ \mathbf{R}_{\mathbf{k}}^{3\pi/2}\}.$$

This collection of linear transformations has two important properties: first, observe that the successive application of any two symmetries yields a third symmetry, i.e. if $\mathbf{P}_1$ and $\mathbf{P}_2$ are in $\mathcal{G}_{\text{square}}$, then so is their product $\mathbf{P}_1\mathbf{P}_2$. For example, $\mathbf{R}_{\mathbf{k}}^{\pi}\mathbf{R}_{\mathbf{k}}^{\pi/2} = \mathbf{R}_{\mathbf{k}}^{3\pi/2}$, $\mathbf{R}_{\mathbf{k}}^{\pi/2}\mathbf{R}_{\mathbf{k}}^{3\pi/2} = \mathbf{I}$, $\mathbf{R}_{\mathbf{k}}^{3\pi/2}\mathbf{R}_{\mathbf{k}}^{3\pi/2} = \mathbf{R}_{\mathbf{k}}^{\pi}$ etc. Second, observe that if $\mathbf{P}$ is any member of $\mathcal{G}_{\text{square}}$, then so is its inverse $\mathbf{P}^{-1}$. For example $(\mathbf{R}_{\mathbf{k}}^{\pi})^{-1} = \mathbf{R}_{\mathbf{k}}^{\pi}$, $(\mathbf{R}_{\mathbf{k}}^{3\pi/2})^{-1} = \mathbf{R}_{\mathbf{k}}^{\pi/2}$ etc. As we shall see in Section 4.4, these two properties endow the set $\mathcal{G}_{\text{square}}$ with a certain special structure.

Next consider the rotation $\mathbf{R}_{\mathbf{k}}^{\pi/2}$ and observe that every element of the set $\mathcal{G}_{\text{square}}$ can be represented in the form $(\mathbf{R}_{\mathbf{k}}^{\pi/2})^n$ for the integer choices $n = 0, 1, 2, 3$. Therefore we can say that the set $\mathcal{G}_{\text{square}}$ is "generated" by the element $\mathbf{R}_{\mathbf{k}}^{\pi/2}$.

Finally observe that

$$\mathcal{G}'_{\text{square}} = \{\mathbf{I}, \mathbf{R}^{\pi}\}$$

is a subset of $\mathcal{G}_{\text{square}}$ *and that it too* has the properties that if $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{G}'_{\text{square}}$ then their product $\mathbf{P}_1\mathbf{P}_2$ is also in $\mathcal{G}'_{\text{square}}$; and if $\mathbf{P} \in \mathcal{G}'_{\text{square}}$ so is its inverse $\mathbf{P}^{-1}$.

We shall generalize all of this in Section 4.4.

4.2 An example in three-dimensions.

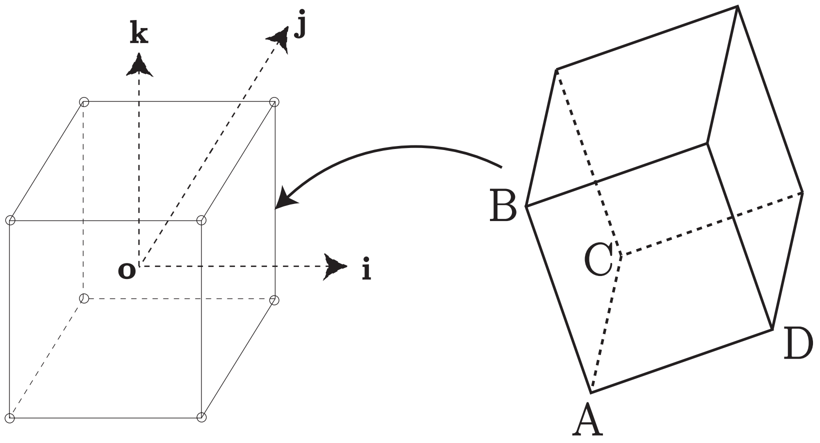


Figure 4.2: Mapping a cube into itself.

Before considering some general theory, it is useful to consider the three-dimensional version of the previous problem. Consider a cube whose center is at the origin **o** and whose edges are parallel to the orthonormal vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, and consider mappings that carry the cube into itself. Consider a vertex A, and its three adjacent vertices B, C, D. The vertex A can be placed in one of 8 positions. Once the location of A has been determined, the vertex B can be placed in one of 3 positions. And once the locations of A and B have been fixed, the vertex C can be placed in one of 2 positions (allowing for reflections or in just one position if only rotations are permitted). Once the vertices A, B and C have been placed, the locations of the remaining vertices are fixed. Thus there are a total of $8 \times 3 \times 2 = 48$ symmetry preserving transformations of the cube, 24 of which are rotations and 24 of which are reflections.

First, consider the 24 rotations. In order to determine these rotations we again (a) identify all axes of rotational symmetry and then (b) determine the number of distinct rotations about each such axis. In the present case we see that, in addition to the identify transformation $\mathbf{I}$ itself, we have the following rotational transformations that preserve symmetry:

1. There are 3 axes that join the center of one face of the cube to the center of the opposite face of the cube which we can take to be $\mathbf{i}, \mathbf{j}, \mathbf{k}$, (which in materials science are called the $\{100\}$ directions); and $90^o, 180^o$ and 270^o rotations about each of these axes maps the cube back into the cube. Thus the following $3 \times 3 = 9$ distinct rotations are symmetry transformations:

$$\mathbf{R}_{\mathbf{i}}^{\pi/2},\ \mathbf{R}_{\mathbf{i}}^{\pi},\ \mathbf{R}_{\mathbf{i}}^{3\pi/2},\ \mathbf{R}_{\mathbf{j}}^{\pi/2},\ \mathbf{R}_{\mathbf{j}}^{\pi},\ \mathbf{R}_{\mathbf{j}}^{3\pi/2},\ \mathbf{R}_{\mathbf{k}}^{\pi/2},\ \mathbf{R}_{\mathbf{k}}^{\pi},\ \mathbf{R}_{\mathbf{k}}^{3\pi/2}$$

2. There are 4 axes that join one vertex of the cube to the diagonally opposite vertex of the cube which we can take to be $\mathbf{i}+\mathbf{j}+\mathbf{k}, \mathbf{i}-\mathbf{j}+\mathbf{k}, \mathbf{i}+\mathbf{j}-\mathbf{k}, \mathbf{i}-\mathbf{j}-\mathbf{k}$, (which in materials science are called the $\{111\}$ directions); and 120^o and 240^o rotations about each of these axes maps the cube back into the cube. Thus the following $4 \times 2 = 8$ distinct rotations are symmetry transformations:

$$\mathbf{R}_{\mathbf{i}+\mathbf{j}+\mathbf{k}}^{2\pi/3},\ \mathbf{R}_{\mathbf{i}+\mathbf{j}+\mathbf{k}}^{4\pi/3},\ \mathbf{R}_{\mathbf{i}-\mathbf{j}+\mathbf{k}}^{2\pi/3},\ \mathbf{R}_{\mathbf{i}-\mathbf{j}+\mathbf{k}}^{4\pi/3},\ \mathbf{R}_{\mathbf{i}+\mathbf{j}-\mathbf{k}}^{2\pi/3},\ \mathbf{R}_{\mathbf{i}+\mathbf{j}-\mathbf{k}}^{4\pi/3},\ \mathbf{R}_{\mathbf{i}-\mathbf{j}-\mathbf{k}}^{2\pi/3},\ \mathbf{R}_{\mathbf{i}-\mathbf{j}-\mathbf{k}}^{4\pi/3}.$$

3. Finally, there are 6 axes that join the center of one edge of the cube to the center of the diagonally opposite edge of the cube which we can take to be $\mathbf{i}+\mathbf{j}, \mathbf{i}-\mathbf{j}, \mathbf{i}+\mathbf{k}, \mathbf{i}-\mathbf{k}, \mathbf{j}+\mathbf{k}, \mathbf{j}-\mathbf{k}$ (which in materials science are called the $\{110\}$ directions); and

180^o rotations about each of these axes maps the cube back into the cube. Thus the following $6 \times 1 = 6$ distinct rotations are symmetry transformations:

$$\mathbf{R}^{\pi}_{\mathbf{i}+\mathbf{j}},\ \mathbf{R}^{\pi}_{\mathbf{i}-\mathbf{j}},\ \mathbf{R}^{\pi}_{\mathbf{i}+\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{i}-\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{j}+\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{j}-\mathbf{k}}.$$

Let $\mathcal{G}_{\text{cube}}$ denote the collection of these 24 symmetry preserving rotations:

$$\begin{aligned}
\mathcal{G}_{\text{cube}} \quad = \quad & \{\mathbf{I}, \\
& \mathbf{R}^{\pi/2}_{\mathbf{i}},\ \mathbf{R}^{\pi}_{\mathbf{i}},\ \mathbf{R}^{3\pi/2}_{\mathbf{i}},\ \mathbf{R}^{\pi/2}_{\mathbf{j}},\ \mathbf{R}^{\pi}_{\mathbf{j}},\ \mathbf{R}^{3\pi/2}_{\mathbf{j}},\ \mathbf{R}^{\pi/2}_{\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{k}},\ \mathbf{R}^{3\pi/2}_{\mathbf{k}} \\
& \mathbf{R}^{2\pi/3}_{\mathbf{i}+\mathbf{j}+\mathbf{k}},\ \mathbf{R}^{4\pi/3}_{\mathbf{i}+\mathbf{j}+\mathbf{k}},\ \mathbf{R}^{2\pi/3}_{\mathbf{i}-\mathbf{j}+\mathbf{k}},\ \mathbf{R}^{4\pi/3}_{\mathbf{i}-\mathbf{j}+\mathbf{k}},\ \mathbf{R}^{2\pi/3}_{\mathbf{i}+\mathbf{j}-\mathbf{k}},\ \mathbf{R}^{4\pi/3}_{\mathbf{i}+\mathbf{j}-\mathbf{k}},\ \mathbf{R}^{2\pi/3}_{\mathbf{i}-\mathbf{j}-\mathbf{k}},\ \mathbf{R}^{4\pi/3}_{\mathbf{i}-\mathbf{j}-\mathbf{k}}, \\
& \mathbf{R}^{\pi}_{\mathbf{i}+\mathbf{j}},\ \mathbf{R}^{\pi}_{\mathbf{i}-\mathbf{j}},\ \mathbf{R}^{\pi}_{\mathbf{i}+\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{i}-\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{j}+\mathbf{k}},\ \mathbf{R}^{\pi}_{\mathbf{j}-\mathbf{k}}\ \}.
\end{aligned} \tag{4.1}$$

If one considers rotations and reflections, then there are 48 elements in this set, where the 24 reflections are obtained by multiplying each rotation by $-\mathbf{I}$. (It is important to remark that this just happens to be true for the cube, but is not generally true. In general, if $\mathbf{R}$ is a rotational symmetry of an object then $-\mathbf{R}$ is, of course, a reflection, but it need not describe a reflectional symmetry of the object; e.g. see the example of the tetrahedron discussed later.)

The collection of linear transformations $\mathcal{G}_{\text{cube}}$ has two important properties that one can verify: (i) if $\mathbf{P}_1$ and $\mathbf{P}_2 \in \mathcal{G}_{\text{cube}}$, then their product $\mathbf{P}_1\mathbf{P}_2$ is also in $\mathcal{G}_{\text{cube}}$, and (ii) if $\mathbf{P} \in \mathcal{G}_{\text{cube}}$, then so does its inverse $\mathbf{P}^{-1}$.

Next, one can verify that every element of the set $\mathcal{G}_{\text{cube}}$ can be represented in the form $(\mathbf{R}^{\pi/2}_{\mathbf{i}})^p(\mathbf{R}^{\pi/2}_{\mathbf{j}})^q(\mathbf{R}^{\pi/2}_{\mathbf{k}})^r$ for integer choices of p, q, r. For example the rotation $\mathbf{R}^{2\pi/3}_{\mathbf{i}+\mathbf{j}+\mathbf{k}}$ (about a $\{111\}$ axis) and the rotation $\mathbf{R}^{\pi}_{\mathbf{i}+\mathbf{k}}$ (about a $\{110\}$ axis) can be represented as

$$\mathbf{R}^{2\pi/3}_{\mathbf{i}+\mathbf{j}+\mathbf{k}} = \left(\mathbf{R}^{\pi/2}_{\mathbf{k}}\right)^{-1}\left(\mathbf{R}^{\pi/2}_{\mathbf{j}}\right)^{-1}, \qquad \mathbf{R}^{\pi}_{\mathbf{i}+\mathbf{k}} = \left(\mathbf{R}^{\pi/2}_{\mathbf{j}}\right)^{-1}\left(\mathbf{R}^{\pi/2}_{\mathbf{k}}\right)^{2}.$$

(One way in which to verify this is to use the representation of a rotation tensor determined in Example 2.18.) Therefore we can say that the set $\mathcal{G}_{\text{cube}}$ is "generated" by the three elements $\mathbf{R}^{\pi/2}_{\mathbf{i}}, \mathbf{R}^{\pi/2}_{\mathbf{j}}$ and $\mathbf{R}^{\pi/2}_{\mathbf{k}}$.

4.3 Lattices.

A geometric structure of particular interest in solid mechanics is a lattice and we now make a few observations on the symmetry of lattices. The simplest lattice, a Bravais lattice $\mathcal{L}\{\mathbf{o}; \boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3\}$, is an infinite set of periodically arranged points in space generated by the translation of a single point $\mathbf{o}$ through three linearly independent lattice vectors $\{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3\}$:

$$\mathcal{L}\{\mathbf{o}; \boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3\} = \{\mathbf{x} \mid \mathbf{x} = \mathbf{o} + \sum_{n=1}^{3} n_i \boldsymbol{\ell}_i, \ n_i \in \mathbb{Z} \ \} \tag{4.2}$$

where $\mathbb{Z}$ is the set of integers. Figure 4.3 shows a two-dimensional square lattice and one possible set of lattice vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2$. (It is clear from the figure that different sets of lattice vectors can correspond to the same lattice.)

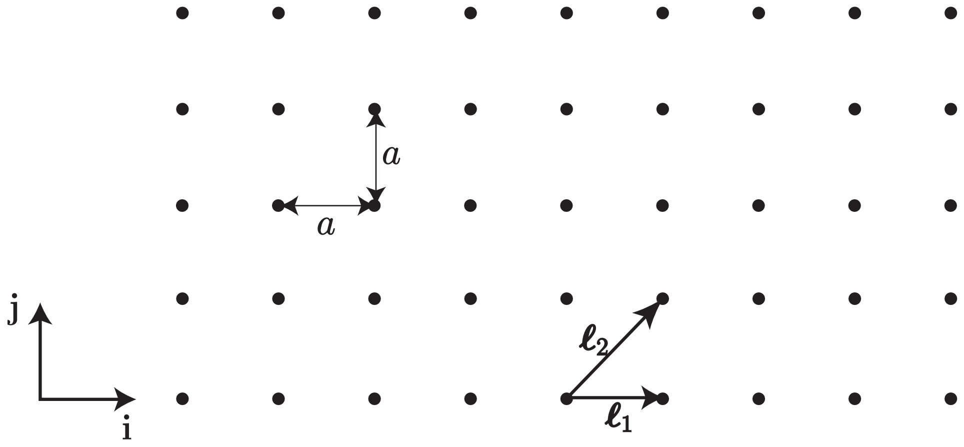


Figure 4.3: A two-dimensional square lattice with lattice vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2$.

It can be shown that a linear transformation $\mathbf{P}$ maps a lattice back into itself if and only if

$$\mathbf{P}\boldsymbol{\ell}_i = \sum_{j=1}^{3} \ M_{ij} \boldsymbol{\ell}_j \tag{4.3}$$

for some 3×3 matrix $[M]$ whose elements M_{ij} are integers and where $\det[M] = 1$. Given a lattice, let $\mathcal{G}_{\text{lattice}}$ be the set of all linear transformations $\mathbf{P}$ that map the lattice back into itself. One can show that if $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{G}_{\text{lattice}}$ then their product $\mathbf{P}_1\mathbf{P}_2$ is also in $\mathcal{G}_{\text{lattice}}$; and if $\mathbf{P} \in \mathcal{G}_{\text{lattice}}$ so is its inverse $\mathbf{P}^{-1}$. The set $\mathcal{G}_{\text{lattice}}$ is called the *symmetry group* of the lattice;

and the set of *rotations* in $\mathcal{G}_{\text{lattice}}$ is known as the *point group* of the lattice. For example the point group of a simple cubic lattice[1] is the set $\mathcal{G}_{\text{cube}}$ of 24 rotations given in (4.1).

4.4 Groups of Linear Transformations.

A collection $\mathcal{G}$ of non-singular linear transformations is said to be a **group of linear transformations** if it possesses the following two properties:

$$(i) \quad \text{if } \mathbf{P}_1 \in \mathcal{G} \text{ and } \mathbf{P}_2 \in \mathcal{G} \text{ then } \mathbf{P}_1\mathbf{P}_2 \in \mathcal{G},$$
$$(ii) \quad \text{if } \mathbf{P} \in \mathcal{G} \text{ then } \mathbf{P}^{-1} \in \mathcal{G}.$$

Note from this that the identity transformation $\mathbf{I}$ is necessarily a member of every group $\mathcal{G}$.

Clearly the three sets $\mathcal{G}_{\text{square}}, \mathcal{G}_{\text{cube}}$ and $\mathcal{G}_{\text{lattice}}$ encountered in the previous sections are groups. One can show that each of the following sets of linear transformations forms a group:

- the set of all orthogonal linear transformations;
- the set of all proper orthogonal linear transformations;
- the set of all unimodular linear transformations[2] (i.e. linear transformations with determinant equal to ±1); and
- the set of all proper unimodular linear transformations (i.e. linear transformations with determinant equal to +1).

The **generators** of a group $\mathcal{G}$ are those elements $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n$ which, when they and their inverses are multiplied among themselves in various combinations yield all the elements of the group. Generators of the groups $\mathcal{G}_{\text{square}}$ and $\mathcal{G}_{\text{cube}}$ were given previously.

In general, a collection of linear transformations $\mathcal{G}'$ is said to be a **subgroup** of a group $\mathcal{G}$ if

$$(i) \quad \mathcal{G}' \subset \mathcal{G} \text{ and}$$
$$(ii) \quad \mathcal{G}' \text{ is itself a group.}$$

[1] There are seven different *types of symmetry* that arise in Bravais lattices, viz. triclinic, monoclinic, orthorhombic, tetragonal, cubic, trigonal and hexagonal. Because, for example, a cubic lattice can be body-centered or face-centered, and so on, the number of different *types of lattices* is greater than seven.

[2] While the determinant of an orthogonal tensor is ±1 the converse is not necessarily true. There are unimodular tensors, e.g. $\mathbf{P} = \mathbf{I} + \alpha\mathbf{i} \otimes \mathbf{j}$, that are not orthogonal. Thus the unimodular group is not equivalent to the orthogonal group.

One can readily show that the group of proper orthogonal linear transformations is a subgroup of the group of orthogonal linear transformations, which in turn is a subgroup of the group of unimodular linear transformations. In our first example, $\mathcal{G}'_{\text{square}}$ is a subgroup of $\mathcal{G}_{\text{square}}$.

It should be mentioned that the general theory of groups deals with collections of elements (together with certain "rules" including "multiplication") where the elements *need not be* linear transformations. For example the set of all integers $\mathbb{Z}$ with "multiplication" defined as the addition of numbers, the identity taken to be zero, and the inverse of x taken to be $-x$ is a group. Similarly the set of all matrices of the form

$$\begin{pmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{pmatrix} \qquad \text{where } -\infty < x < \infty,$$

with "multiplication" defined as matrix multiplication, the identity being the identity matrix, and the inverse being

$$\begin{pmatrix} \cosh(-x) & \sinh(-x) \\ \sinh(-x) & \cosh(-x) \end{pmatrix},$$

can be shown to be a group. However, our discussion in these notes is limited to groups of linear transformations.

4.5 Symmetry of a scalar-valued function of symmetric positive-definite tensors.

When we discuss the constitutive behavior of a material in Volume 2, we will encounter a scalar-valued function $\psi(\mathbf{C})$ defined for all symmetric positive definite tensors $\mathbf{C}$. (This represents the energy in the material and characterizes its mechanical response). The symmetry of the material will be characterized by a set $\mathcal{G}$ of non-singular tensors $\mathbf{P}$ which has the property that, for each $\mathbf{P} \in \mathcal{G}$,

$$\psi(\mathbf{C}) = \psi(\mathbf{P}^T\mathbf{C}\mathbf{P}) \qquad \text{for all symmetric positive} - \text{definite } \mathbf{C}. \tag{4.4}$$

It can be readily shown that this set of tensors $\mathcal{G}$ is a group. To see this, first let $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{G}$ so that

$$\psi(\mathbf{C}) = \psi(\mathbf{P}_1^T\mathbf{C}\mathbf{P}_1), \qquad \psi(\mathbf{C}) = \psi(\mathbf{P}_2^T\mathbf{C}\mathbf{P}_2), \tag{4.5}$$

for all symmetric positive-definite $\mathbf{C}$. Then $\psi((\mathbf{P}_1\mathbf{P}_2)^T\mathbf{C}\mathbf{P}_1\mathbf{P}_2) = \psi(\mathbf{P}_2^T(\mathbf{P}_1^T\mathbf{C}\mathbf{P}_1)\mathbf{P}_2) = \psi(\mathbf{P}_1^T\mathbf{C}\mathbf{P}_1) = \psi(\mathbf{C})$ where we have used $(4.5)_2$ and $(4.5)_1$ in the penultimate and ultimate steps respectively. Thus if $\mathbf{P}_1$ and $\mathbf{P}_2$ are in $\mathcal{G}$, then so is $\mathbf{P}_1\mathbf{P}_2$. Next, suppose that $\mathbf{P} \in \mathcal{G}$. Since $\mathbf{P}$ is non-singular, the equation $\mathbf{S} = \mathbf{P}^T\mathbf{C}\mathbf{P}$ provides a one-to-one relation between symmetric positive definite tensors $\mathbf{C}$ and $\mathbf{S}$. Thus, since (4.4) holds for all symmetric positive-definite $\mathbf{C}$, it also holds for all symmetric positive-definite linear transformations $\mathbf{S} = \mathbf{P}^T\mathbf{C}\mathbf{P}$. Substituting this into (4.4) gives $\psi(\mathbf{S}) = \psi(\mathbf{P}^{-T}\mathbf{S}\mathbf{P}^{-1})$ for all symmetric positive-definite $\mathbf{S}$; and so $\mathbf{P}^{-1}$ is also in $\mathcal{G}$. Thus the set $\mathcal{G}$ of nonsingular tensors obeying (4.4) is a group; we shall refer to it as the **symmetry group** of ψ.

Observe from (4.4) that the symmetry group of ψ contains the elements $\mathbf{I}$ and $-\mathbf{I}$, and as a consequence, if $\mathbf{P} \in \mathcal{G}$ then $-\mathbf{P} \in \mathcal{G}$ also.

To examine an explicit example, consider the function

$$\psi(\mathbf{C}) = \widehat{\psi}\Big(\det \mathbf{C}\Big). \tag{4.6}$$

It is seen trivially that for this $\widehat{\psi}$, equation (4.4) holds if and only if $\det \mathbf{P} = \pm 1$. Thus the symmetry group of this ψ consists of all unimodular tensors (i.e. tensors with determinant equal to ±1).

As a second example consider the function

$$\psi(\mathbf{C}) = \widehat{\psi}\Big(\mathbf{C}\mathbf{n} \cdot \mathbf{n}\Big) \tag{4.7}$$

where $\mathbf{n}$ is a given fixed unit vector. Let $\mathbf{Q}_{\rm n}$ be a rotation about the axis $\mathbf{n}$ through an arbitrary angle. Then since $\mathbf{n}$ is the axis of $\mathbf{Q}_{\rm n}$ we know that $\mathbf{Q}_{\rm n}\mathbf{n} = \mathbf{n}$. Therefore

$$\psi(\mathbf{Q}_{\rm n}^T\mathbf{C}\mathbf{Q}_{\rm n}) = \widehat{\psi}\Big(\mathbf{Q}_{\rm n}^T\mathbf{C}\mathbf{Q}_{\rm n}\mathbf{n} \cdot \mathbf{n}\Big) = \widehat{\psi}\Big(\mathbf{C}\mathbf{Q}_{\rm n}\mathbf{n} \cdot \mathbf{Q}_{\rm n}\mathbf{n}\Big) = \widehat{\psi}\Big(\mathbf{C}\mathbf{n} \cdot \mathbf{n}\Big) = \psi(\mathbf{C}). \tag{4.8}$$

The symmetry group of the function (4.7) therefore contains the set of all rotations about $\mathbf{n}$. (Are there any other tensors in $\mathcal{G}$?)

The following result will be useful in Volume 2. Let $\mathbf{H}$ be some fixed nonsingular linear transformation, and consider two functions $\psi_1(\mathbf{C})$ and $\psi_2(\mathbf{C})$, each defined for all symmetric positive-definite tensors $\mathbf{C}$. Suppose that ψ_1 and ψ_2 are related by

$$\psi_2(\mathbf{C}) = \psi_1(\mathbf{H}^T\mathbf{C}\mathbf{H}) \quad \text{for all symmetric positive} - \text{definite tensors } \mathbf{C}. \tag{4.9}$$

If $\mathcal{G}_1$ and $\mathcal{G}_2$ are the symmetry groups of ψ_1 and ψ_2 respectively, then it can be shown that

$$\mathcal{G}_2 = \mathbf{H}\mathcal{G}_1\mathbf{H}^{-1} \tag{4.10}$$

in the sense that a tensor $\mathbf{P} \in \mathcal{G}_1$ if and only if the tensor $\mathbf{H}\mathbf{P}\mathbf{H}^{-1} \in \mathcal{G}_2$. As a special case of this, if $\mathbf{H}$ is a spherical tensor, i.e. if $\mathbf{H} = \alpha\mathbf{I}$, then $\mathcal{G}_1 = \mathcal{G}_2$.

Next, note that any nonsingular tensor $\mathbf{P}$ can be written as the product of a spherical tensor $\alpha\mathbf{I}$ and a unimodular tensor $\mathbf{T}$ as $\mathbf{P} = (\alpha\mathbf{I})\mathbf{T}$ provided that we take $\alpha = (|\det\mathbf{P}|)^{1/3}$ since then $\det\mathbf{T} = \pm 1$. This, together with the special case of the result noted in the preceding paragraph provides a hint of why we might want to limit attention to *unimodular* tensors rather than consider all nonsingular tensors in our discussion of symmetry.

This motivates the following slight modification to our original notion of symmetry of a function $\psi(\mathbf{C})$. We characterize the symmetry of ψ by the set $\mathcal{G}$ of *unimodular* tensors $\mathbf{P}$ which have the property that, for each $\mathbf{P} \in \mathcal{G}$,

$$\psi(\mathbf{C}) = \psi(\mathbf{P}^T\mathbf{C}\mathbf{P}) \qquad \text{for all symmetric positive} - \text{definite } \mathbf{C}. \tag{4.11}$$

It can be readily shown that this set of tensors $\mathcal{G}$ is also a group, necessarily a subgroup of the unimodular group.

A function ψ is said to be *isotropic* if its symmetry group $\mathcal{G}$ contains all orthogonal tensors. Thus for an isotropic function ψ,

$$\psi(\mathbf{C}) = \psi(\mathbf{P}^T\mathbf{C}\mathbf{P}) \tag{4.12}$$

for all symmetric positive-definite $\mathbf{C}$ and all orthogonal $\mathbf{P}$. From a theorem in algebra it follows that an isotropic function ψ depends on $\mathbf{C}$ only through its principal scalar invariants defined previously in (3.38), i.e. that there exists a function $\widehat{\psi}$ such that

$$\psi(\mathbf{C}) = \widehat{\psi}\Big(I_1(\mathbf{C}), I_2(\mathbf{C}), I_3(\mathbf{C})\Big) \tag{4.13}$$

where

$$\left.\begin{aligned} I_1(\mathbf{C}) &= \text{trace } \mathbf{C} \\ I_2(\mathbf{C}) &= 1/2\,[(\text{trace } \mathbf{C})^2 - \text{trace } (\mathbf{C}^2)]\,, \\ I_3(\mathbf{C}) &= \det \mathbf{C}. \end{aligned}\right\} \tag{4.14}$$

As a second example consider "cubic symmetry" where the symmetry group $\mathcal{G}$ coincides with the set of 24 rotations $\mathcal{G}_{\text{cube}}$ given in (4.1) plus the corresponding reflections obtained by multiplying these rotations by $-\mathbf{I}$. As noted previously, this group is generated by $\mathbf{R}^{\pi/2}_{\mathbf{i}}, \mathbf{R}^{\pi/2}_{\mathbf{j}}, \mathbf{R}^{\pi/2}_{\mathbf{k}}$ and $-\mathbf{I}$, and contains 24 rotations and 24 reflections. Then, according to a theorem in algebra (see pg 312 of Truesdell and Noll),

$$\psi(\mathbf{C}) = \widehat{\psi}\Big(i_1(\mathbf{C}), i_2(\mathbf{C}), i_3(\mathbf{C}), i_4(\mathbf{C}), i_5(\mathbf{C}), i_6(\mathbf{C}), i_7(\mathbf{C}), i_8(\mathbf{C}), i_9(\mathbf{C})\Big) \tag{4.15}$$

where

$$\left.\begin{aligned} i_1(\mathbf{C}) &= C_{11} + C_{22} + C_{33}, \\ i_2(\mathbf{C}) &= C_{22}C_{33} + C_{33}C_{11} + C_{11}C_{22} \\ i_3(\mathbf{C}) &= C_{11}C_{22}C_{33} \\ i_4(\mathbf{C}) &= C_{23}^2 + C_{31}^2 + C_{12}^2 \\ i_5(\mathbf{C}) &= C_{31}^2C_{32}^2 + C_{12}^2C_{23}^2 + C_{23}^2C_{31}^2 \\ i_6(\mathbf{C}) &= C_{23}C_{31}C_{12} \\ i_7(\mathbf{C}) &= C_{22}C_{12}^2 + C_{33}C_{31}^2 + C_{33}C_{23}^2 + C_{11}C_{12}^2 + C_{11}C_{31}^2 + C_{22}C_{23}^2 \\ i_8(\mathbf{C}) &= C_{11}C_{31}^2C_{12}^2 + C_{22}C_{12}^2C_{23}^2 + C_{33}C_{23}^2C_{31}^2 \\ i_9(\mathbf{C}) &= C_{23}^2C_{22}C_{33} + C_{31}^2C_{33}C_{11} + C_{12}^2C_{11}C_{22} \end{aligned}\right\} \tag{4.16}$$

If $\mathcal{G}$ contains $\mathbf{I}$ and all rotations $\mathbf{R}^{\phi}_{\mathbf{n}}$, $0 < \phi < 2\pi$, through all angles ϕ about a fixed axis $\mathbf{n}$, the corresponding symmetry is called *transverse isotropy.*

If $\mathcal{G}$ includes the three elements $-\mathbf{R}^{\pi}_{\mathbf{i}}$, $-\mathbf{R}^{\pi}_{\mathbf{j}}$, $-\mathbf{R}^{\pi}_{\mathbf{k}}$ which represent reflections in the planes normal to $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, the symmetry is called *orthotropy.*

4.6 Worked Examples.

Example 4.1: Characterize the set $\mathcal{H}_{\text{square}}$ of linear transformations that map a square back into a square, including both rotations and reflections.

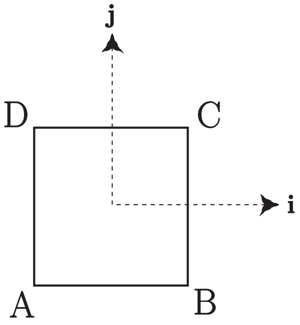


Figure 4.4: Mapping a square into itself.

Solution: We return to the problem describe in Section 4.1 and now consider the set rotations and reflections $\mathcal{H}_{\text{square}}$ that map the square back into itself. The set of rotations that do this were determined earlier and they are

$$\mathcal{G}_{\text{square}} = \{\mathbf{I},\ \mathbf{R}_{\mathbf{k}}^{\pi/2},\ \mathbf{R}_{\mathbf{k}}^{\pi},\ \mathbf{R}_{\mathbf{k}}^{3\pi/2}\}.$$

As the 4 reflectional symmetries we can pick

$$\begin{aligned}
\mathbf{H} &= \text{reflection in the horizontal axis } \mathbf{i},\\
\mathbf{V} &= \text{reflection in the vertical axis } \mathbf{j},\\
\mathbf{D} &= \text{reflection in the diagonal with positive slope } \mathbf{i}+\mathbf{j},\\
\mathbf{D}' &= \text{reflection in the diagonal with negative slope } -\mathbf{i}+\mathbf{j},
\end{aligned}$$

and so

$$\mathcal{H}_{\text{square}} = \left\{\mathbf{I},\ \mathbf{R}_{\mathbf{k}}^{\pi/2}, \mathbf{R}_{\mathbf{k}}^{\pi}, \mathbf{R}_{\mathbf{k}}^{3\pi/2}, \mathbf{H}, \mathbf{V}, \mathbf{D}, \mathbf{D}'\right\} . \tag{i}$$

One can verify that $\mathcal{H}_{\text{square}}$ is a group since it possesses the property that if $\mathbf{P}_1$ and $\mathbf{P}_2$ are two transformations in $\mathcal{G}$, then so is their product $\mathbf{P}_1\mathbf{P}_2$; e.g. $\mathbf{D}' = \mathbf{R}_{\mathbf{k}}^{3\pi/2}\mathbf{H},\ \mathbf{D} = \mathbf{H}\mathbf{R}_{\mathbf{k}}^{3\pi/2}$ etc. And if $\mathbf{P}$ is any member of $\mathcal{G}$, then so is its inverse; e.g. $\mathbf{H}^{-1} = \mathbf{H}$ etc.

Example 4.2: Find the generators of $\mathcal{H}_{\text{square}}$ and all subgroups of $\mathcal{H}_{\text{square}}$.

Solution: All elements of $\mathcal{H}_{\text{square}}$ can be represented in the form $(\mathbf{R}_{\mathbf{k}}^{\pi/2})^i\mathbf{H}^j$ for integer choices of $i = 0, 1, 2, 3$ and $j = 0, 1$:

$$\mathbf{R}_{\mathbf{k}}^{\pi} = (\mathbf{R}_{\mathbf{k}}^{\pi/2})^2, \quad \mathbf{R}_{\mathbf{k}}^{3\pi/2} = (\mathbf{R}_{\mathbf{k}}^{\pi/2})^3, \quad \mathbf{I} = (\mathbf{R}_{\mathbf{k}}^{\pi/2})^4.$$

$$\mathbf{D}' = (\mathbf{R}_{\mathbf{k}}^{\pi/2})^3\mathbf{H}, \quad \mathbf{V} = (\mathbf{R}_{\mathbf{k}}^{\pi/2})^2\mathbf{H}, \quad \mathbf{D} = \mathbf{R}_{\mathbf{k}}^{\pi/2}\mathbf{H}.$$

Therefore the group $\mathcal{H}_{\text{square}}$ is generated by the two elements $\mathbf{H}$ and $\mathbf{R}^{\pi/2}$.

One can verify that the following 8 collections of linear transformations are subgroups of $\mathcal{H}_{\text{square}}$:

$$\left\{\mathbf{I}, \mathbf{R}^{\pi/2}, \mathbf{R}^{\pi}, \mathbf{R}^{3\pi/2}\right\}, \quad \{\mathbf{I}, \mathbf{D}, \mathbf{D}', \mathbf{R}^{\pi}\}, \quad \{\mathbf{I}, \mathbf{H}, \mathbf{V}, \mathbf{R}^{\pi}\}, \quad \{\mathbf{I}, \mathbf{R}^{\pi}\},$$

$$\{\mathbf{I}, \mathbf{D}\}, \quad \{\mathbf{I}, \mathbf{D}'\}, \quad \{\mathbf{I}, \mathbf{H}\}, \quad \{\mathbf{I}, \mathbf{V}\},$$

Geometrically, each of these subgroups leaves some aspect of the square invariant. The first leaves the face invariant, the second leaves a diagonal invariant, the third leaves the axis invariant, the fourth leaves an axis and a diagonal invariant etc. There are *no* other subgroups of $\mathcal{H}_{\text{square}}$.

Example 4.3: Characterize the rotational symmetry of a regular tetrahedron.

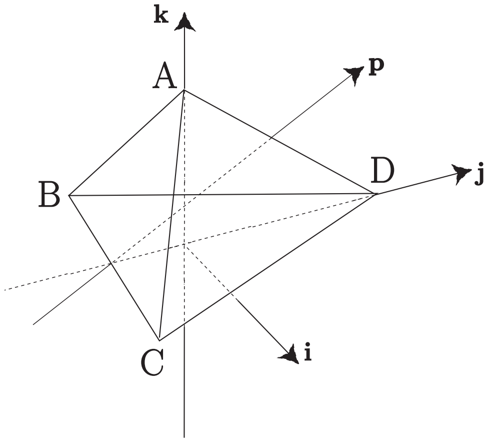


Figure 4.5: A regular tetrahedron ABCD, three orthonormal vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ and a unit vector $\mathbf{p}$. The axis $\mathbf{k}$ passes through the vertex A and the centroid of the opposite face BCD, while the unit vector $\mathbf{p}$ passes through the center of the edge AD and the center of the opposite edge BC.

Solution:

1. There are 4 axes like $\mathbf{k}$ in the figure that pass through a vertex of the tetrahedron and the centroid of the opposite face; and right-handed rotations of 120^o and 240^o about each of these axes maps the tetrahedron back onto itself. Thus these $4 \times 2 = 8$ distinct rotations – of the form $\mathbf{R}_{\mathbf{k}}^{2\pi/3}, \mathbf{R}_{\mathbf{k}}^{4\pi/3}$, etc. – are symmetry transformations of the tetrahedron.
2. There are three axes like $\mathbf{p}$ shown in the figure that pass through the mid-points of a pair of opposite edges; and a right-handed rotation through 180^o about each of these axes maps the tetrahedron back onto itself. Thus these $3 \times 1 = 3$ distinct rotations – of the form $\mathbf{R}_{\mathbf{p}}^{\pi}$, etc. – are symmetry transformations of the tetrahedron.

The group $\mathcal{G}_{\text{tetrahedron}}$ of rotational symmetries of a tetrahedron therefore consists of these 11 rotations plus the identity transformation $\mathbf{I}$.

Example 4.4: Are all symmetry preserving linear transformations necessarily either rotations or reflections?

Solution: We began this chapter by considering the symmetry of a square, and examining the different ways in which the square could be mapped back into itself. Now consider the example of a two-dimensional $a \times a$

square lattice, i.e. the set of infinite points

$$\mathcal{L}_{\text{square}} = \{\mathbf{x} \mid \mathbf{x} = n_1 a\mathbf{i} + n_2 a\mathbf{j},\ n_1, n_2 \in \mathbb{Z} \equiv \text{Integers}\} \tag{i}$$

depicted in Figure 4.6, and examine the different ways in which this lattice can be mapped back into itself.

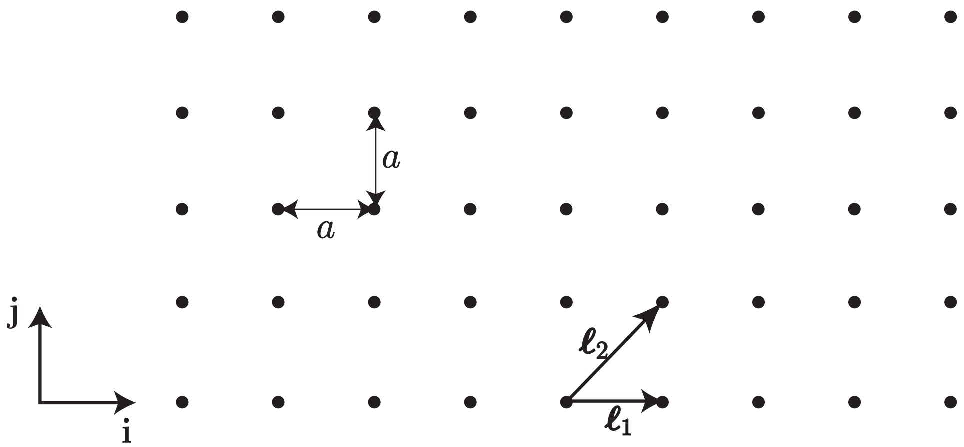


Figure 4.6: A two-dimensional $\ell \times \ell$ square lattice.

We first note that the rotational and reflectional symmetry transformations of a $a \times a$ square are also symmetry transformations for the lattice since they leave the lattice invariant. There are however other transformations, *that are neither rotations nor reflections*, that also leave the lattice invariant. For example, if for every integer n, one rigidly translates the n^{th} row of the lattice by precisely the amount $n\ell$ in the $\mathbf{i}$ direction, one recovers the original lattice. Thus, the "shearing" of the lattice described by the linear transformation

$$\mathbf{P} = \mathbf{I} + a\mathbf{i} \otimes \mathbf{j} \tag{ii}$$

is also a symmetry preserving transformation.

Example 4.5: Show that each of the following sets of linear transformations forms a group: all orthogonal tensors; all proper orthogonal tensors; all unimodular tensors (i.e. tensors with determinant equal to ±1); and all proper unimodular tensors (i.e. tensors with determinant equal to +1).

Example 4.6: Show that the group of proper orthogonal tensors is a subgroup of the group of orthogonal tensors, which in turn is a subgroup of the group of unimodular tensors.

Example 4.7: Suppose that a function $\psi(\mathbf{C})$ is defined for all symmetric positive definite tensors $\mathbf{C}$ and that its symmetry group is the set of all orthogonal tensors. Show that ψ depends on $\mathbf{C}$ only through its principal scalar invariants, i.e. show that there is a function $\widehat{\psi}$ such that

$$\psi(\mathbf{C}) = \widehat{\psi}\Big(I_1(\mathbf{C}), I_2(\mathbf{C}), I_3(\mathbf{C})\Big)$$

where $I_i(\mathbf{C})$, $i = 1, 2, 3$, are the principal scalar invariants of $\mathbf{C}$ defined previously in (3.38).

Solution: We are given that ψ has the property that for all symmetric positive-definite tensors $\mathbf{C}$ and all orthogonal tensors $\mathbf{Q}$

$$\psi(\mathbf{C}) = \psi(\mathbf{Q}^T\mathbf{C}\mathbf{Q}). \tag{i}$$

In order to prove the desired result it is sufficient to show that, if $\mathbf{C}_1$ and $\mathbf{C}_2$ are two symmetric tensors whose principal invariants I_i are the same,

$$I_1(\mathbf{C}_1) = I_1(\mathbf{C}_2), \qquad I_2(\mathbf{C}_1) = I_2(\mathbf{C}_2), \qquad I_3(\mathbf{C}_1) = I_3(\mathbf{C}_2), \tag{ii}$$

then $\psi(\mathbf{C}_1) = \psi(\mathbf{C}_2)$.

Recall that the mapping (3.40) between principal invariants and eigenvalues is one-to-one. It follows from this and (ii) that the eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ are the same. Thus we can write

$$\mathbf{C}_1 = \sum_{i=1}^{3} \lambda_i \mathbf{e}_i^{(1)} \otimes \mathbf{e}_i^{(1)}, \qquad \mathbf{C}_2 = \sum_{i=1}^{3} \lambda_i \mathbf{e}_i^{(2)} \otimes \mathbf{e}_i^{(2)}, \tag{iii}$$

where the two sets of orthonormal vectors $\{\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \mathbf{e}_3^{(1)}\}$ and $\{\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \mathbf{e}_3^{(1)}\}$ are the respective principal bases of $\mathbf{C}_1$ and $\mathbf{C}_2$. Since each set of basis vectors is orthonormal, there is an orthogonal tensor $\mathbf{R}$ that carries $\{\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \mathbf{e}_3^{(1)}\}$ into $\{\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \mathbf{e}_3^{(2)}\}$:

$$\mathbf{R}\mathbf{e}_i^{(1)} = \mathbf{e}_i^{(2)}, \qquad i = 1, 2, 3. \tag{iv}$$

Thus

$$\mathbf{R}^T\left(\sum_{i=1}^{3} \lambda_i \mathbf{e}_i^{(2)} \otimes \mathbf{e}_i^{(2)}\right)\mathbf{R} = \sum_{i=1}^{3} \lambda_i \mathbf{R}^T(\mathbf{e}_i^{(2)} \otimes \mathbf{e}_i^{(2)})\mathbf{R} = \sum_{i=1}^{3} \lambda_i (\mathbf{R}^T\mathbf{e}_i^{(2)}) \otimes (\mathbf{R}^T\mathbf{e}_i^{(2)}) = \sum_{i=1}^{3} \lambda_i (\mathbf{e}_i^{(1)} \otimes (\mathbf{e}_i^{(1)}), \tag{v}$$

and so $\mathbf{R}^T\mathbf{C}_2\mathbf{R} = \mathbf{C}_1$. Therefore $\psi(\mathbf{C}_1) = \psi(\mathbf{R}^T\mathbf{C}_2\mathbf{R}) = \psi(\mathbf{C}_2)$ where in the last step we have used (i). This establishes the desired result.

Example 4.8: Consider a scalar-valued function $f(\mathbf{x})$ that is defined for all vectors $\mathbf{x}$. Let $\mathcal{G}$ be the set of all non-singular linear transformations $\mathbf{P}$ that have the property that for each $\mathbf{P} \in \mathcal{G}$, one has $f(\mathbf{x}) = f(\mathbf{P}\mathbf{x})$ for all vectors $\mathbf{x}$.

i) Show that $\mathcal{G}$ is a group.

ii) Find the most general form of f if $\mathcal{G}$ contains the set of all orthogonal transformations.

Solution:

i) Suppose that $\mathbf{P}_1$ and $\mathbf{P}_2$ are in $\mathcal{G}$, i.e. that

$$\left.\begin{aligned} f(\mathbf{x}) &= f(\mathbf{P}_1\mathbf{x}) \qquad \text{for all vectors } \mathbf{x}, \quad \text{and} \\ f(\mathbf{x}) &= f(\mathbf{P}_2\mathbf{x}) \qquad \text{for all vectors } \mathbf{x}. \end{aligned}\right\} \tag{i}$$

Then

$$f\big((\mathbf{P}_1\mathbf{P}_2)\mathbf{x}\big) = f\big(\mathbf{P}_1(\mathbf{P}_2\mathbf{x})\big) = f(\mathbf{P}_2\mathbf{x}) = f(\mathbf{x})$$

where in the penultimate and ultimate steps we have used (i)$_1$ and (i)$_2$ respectively.

Next, suppose that $\mathbf{P} \in \mathcal{G}$ so that

$$f(\mathbf{x}) = f(\mathbf{P}\mathbf{x}) \qquad \text{for all vectors } \mathbf{x}.$$

Since $\mathbf{P}$ is non-singular we can set $\mathbf{y} = \mathbf{P}\mathbf{x}$ and obtain

$$f(\mathbf{P}^{-1}\mathbf{y}) = f(\mathbf{y}) \qquad \text{for all vectors } \mathbf{y}.$$

It thus follows that $\mathcal{G}$ has the two defining properties of a group.

ii) If $\mathbf{x}_1$ and $\mathbf{x}_2$ are two vectors that have the same length, we will show that $f(\mathbf{x}_1) = f(\mathbf{x}_2)$, whence $f(\mathbf{x})$ depends on $\mathbf{x}$ only through its length $|\mathbf{x}|$, i.e. there exists a function $\widehat{f}$ such that

$$f(\mathbf{x}) = \widehat{f}(|\mathbf{x}|) \qquad \text{for all vectors } \mathbf{x}.$$

If $\mathbf{x}_1$ and $\mathbf{x}_2$ are two vectors that have the same length, there is a rotation tensor $\mathbf{R}$ that carries $\mathbf{x}_2$ to $\mathbf{x}_1$: $\mathbf{R}\mathbf{x}_2 = \mathbf{x}_1$. Therefore

$$f(\mathbf{x}_1) = f(\mathbf{R}\mathbf{x}_2) = f(\mathbf{x}_2),$$

where in the last step we have used the fact that $\mathcal{G}$ contains the set of all orthogonal transformations, i.e. that $f(\mathbf{x}) = f(\mathbf{P}\mathbf{x})$ for all vectors $\mathbf{x}$ and all orthogonal $\mathbf{P}$. This establishes the result claimed above.

Example 4.9: Consider a scalar-valued function $g(\mathbf{C}, \mathbf{m}\otimes\mathbf{m})$ that is defined for all symmetric positive-definite tensors $\mathbf{C}$ and all unit vectors $\mathbf{m}$. Let $\mathcal{G}$ be the set of all non-singular linear transformations $\mathbf{P}$ that have the property that for each $\mathbf{P} \in \mathcal{G}$, one has $g(\mathbf{C}, \mathbf{n}\otimes\mathbf{n}) = g(\mathbf{P}^T\mathbf{C}\mathbf{P}, \mathbf{P}^T(\mathbf{n}\otimes\mathbf{n})\mathbf{P})$ for all symmetric positive-definite tensors $\mathbf{C}$ and *some particular* unit vector $\mathbf{n}$. If $\mathcal{G}$ contains the set of all orthogonal transformations, show that there exists a function $\widehat{g}$ such that

$$g(\mathbf{C}, \mathbf{n}\otimes\mathbf{n}) = \widehat{g}\Big(I_1(\mathbf{C}), I_2(\mathbf{C}), I_3(\mathbf{C}), I_4(\mathbf{C}, \mathbf{n}), I_5(\mathbf{C}, \mathbf{n})\Big)$$

where $I_1(\mathbf{C}), I_2(\mathbf{C}), I_3(\mathbf{C})$ are the three fundamental scalar invariants of $\mathbf{C}$ and

$$I_4(\mathbf{C}, \mathbf{n}) = \mathbf{C}\mathbf{n}\cdot\mathbf{n}, \qquad I_5(\mathbf{C}, \mathbf{n}) = \mathbf{C}^2\mathbf{n}\cdot\mathbf{n}.$$

Remark: Observe that with respect to an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where $\mathbf{e}_3 = \mathbf{n}$ one has $I_4 = C_{33}$ and $I_5 = C_{31}^2 + C_{32}^2 + C_{33}^2$.

Solution: We are told that

$$g(\mathbf{C}, \mathbf{n}\otimes\mathbf{n}) = g(\mathbf{Q}^T\mathbf{C}\mathbf{Q}, \mathbf{Q}^T(\mathbf{n}\otimes\mathbf{n})\mathbf{Q}) \tag{i}$$

for all orthogonal $\mathbf{Q}$ and all symmetric positive definite $\mathbf{C}$. As in Example 4.7, it is sufficient to show that if $\mathbf{C}_1$ and $\mathbf{C}_2$ are two symmetric positive definite linear transformations whose "invariants" $I_i, i = 1, 2, 3, 4, 5$" are the same, i.e.

$$I_1(\mathbf{C}_1) = I_1(\mathbf{C}_2),\ I_2(\mathbf{C}_1) = I_2(\mathbf{C}_2),\ I_3(\mathbf{C}_1) = I_3(\mathbf{C}_2),\ I_4(\mathbf{C}_1, \mathbf{n}) = I_4(\mathbf{C}_2, \mathbf{n}),\ I_5(\mathbf{C}_1, \mathbf{n}) = I_5(\mathbf{C}_2, \mathbf{n}) \tag{ii}$$

then $g(\mathbf{C}_1, \mathbf{n}\otimes\mathbf{n}) = g(\mathbf{C}_2, \mathbf{n}\otimes\mathbf{n})$. From (ii)$_{1,2,3}$ and the analysis in Example 4.7 it follows that there is an orthogonal tensor $\mathbf{R}$ such that $\mathbf{R}^T\mathbf{C}_2\mathbf{R} = \mathbf{C}_1$. It is readily seen from this that $\mathbf{R}^T\mathbf{C}_2^2\mathbf{R} = \mathbf{C}_1^2$ as well. It now follows from this, the fact that $\mathbf{R}$ is orthogonal, (ii)$_{4,5}$ and the definitions of I_4 and I_5 that

$$\mathbf{Rn}\cdot\mathbf{Rn} = \mathbf{n}\cdot\mathbf{n}, \qquad \mathbf{C}_2\mathbf{Rn}\cdot\mathbf{Rn} = \mathbf{C}_2\mathbf{n}\cdot\mathbf{n}, \qquad \mathbf{C}_2^2\mathbf{Rn}\cdot\mathbf{Rn} = \mathbf{C}_2^2\mathbf{n}\cdot\mathbf{n}, \tag{iii}$$

and this must hold for all symmetric positive define $\mathbf{C}_2$. This implies that

$$\mathbf{Rn} = \pm\mathbf{n} \qquad \text{and consequently} \qquad \mathbf{R}^T\mathbf{n} = \pm\mathbf{n},$$

as may be seen, for example, for expressing (iii) in a principal basis of $\mathbf{C}_2$. Consequently

$$g(\mathbf{C}_1, \mathbf{n}\otimes\mathbf{n}) = g(\mathbf{R}^T\mathbf{C}_2\mathbf{R}, (\mathbf{R}^T\mathbf{n})\otimes(\mathbf{R}^T\mathbf{n})) = g(\mathbf{R}^T\mathbf{C}_2\mathbf{R}, \mathbf{R}^T(\mathbf{n}\otimes\mathbf{n})\mathbf{R}) = g(\mathbf{C}_2, \mathbf{n}\otimes\mathbf{n})$$

where we have used (i) in the very last step. This establishes the desired result.

REFERENCES

1. M.A. Armstrong, *Groups and Symmetry*, Springer-Verlag, 1988.
2. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, MacMillan, 1977.
3. C. Truesdell and W. Noll, The nonlinear field theories of mechanics, in *Handbuch der Physik*, Volume III/3, edited by S. Flugge, Springer-Verlag, 1965.
4. A.J.M. Spencer, Theory of invariants, in *Continuum Physics*, Volume I, edited by A.C. Eringen, Academic Press, 1971.

Chapter 5

Calculus of Vector and Tensor Fields

Notation:

α	.....	scalar
$\{a\}$	.....	3×1 column matrix
$\mathbf{a}$	.....	vector
a_i	.....	i^{th} component of the vector $\mathbf{a}$ in some basis; or i^{th} element of the column matrix $\{a\}$
$[A]$	.....	3×3 square matrix
$\mathbf{A}$	.....	second-order tensor (2-tensor)
A_{ij}	.....	i, j component of the 2-tensor $\mathbf{A}$ in some basis; or i, j element of the square matrix $[A]$
$\mathbb{C}$	.....	fourth-order tensor (4-tensor)
$\mathbb{C}_{ijk\ell}$	.....	i, j, k, ℓ component of 4-tensor $\mathbb{C}$ in some basis
$\mathbb{T}_{i_1 i_2 \ldots . i_n}$	.....	$i_1 i_2 \ldots . i_n$ component of n-tensor $\mathbb{T}$ in some basis.

5.1 Notation and definitions.

Let $\mathcal{R}$ be a bounded region of three-dimensional space whose boundary is denoted by $\partial\mathcal{R}$ and let $\mathbf{x}$ denote the position vector of a generic point in $\mathcal{R} + \partial\mathcal{R}$. We shall consider scalar and tensor fields such as $\phi(\mathbf{x}), \mathbf{v}(\mathbf{x}), \mathbf{A}(\mathbf{x})$ and $\mathbb{T}(\mathbf{x})$ defined on $\mathcal{R} + \partial\mathcal{R}$. The region $\mathcal{R} + \partial\mathcal{R}$ and these fields will always be assumed to be sufficiently regular so as to permit the calculations carried out below.

While the subject of the calculus of tensor fields can be dealt with directly, we shall take the more limited approach of working with the components of these fields. The components will always be taken with respect to a single fixed orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Each component of say a vector field $\mathbf{v}(\mathbf{x})$ or a 2-tensor field $\mathbf{A}(\mathbf{x})$ is effectively a scalar-valued function on three-dimensional space, $v_i(x_1, x_2, x_3)$ and $A_{ij}(x_1, x_2, x_3)$, and we can use the well-known operations of classical calculus on such fields such as partial differentiation with respect to x_k.

In order to simplify writing, we shall use the notation that a comma followed by a subscript denotes partial differentiation with respect to the corresponding x-coordinate. Thus, for example, we will write

$$\phi_{,i} = \frac{\partial \phi}{\partial x_i}, \qquad \phi_{,ij} = \frac{\partial^2 \phi}{\partial x_i \partial x_j}, \qquad v_{i,j} = \frac{\partial v_i}{\partial x_j}, \tag{5.1}$$

and so on, where v_i and x_i are the i^{th} components of the vectors $\mathbf{v}$ and $\mathbf{x}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

The **gradient** of a scalar field $\phi(\mathbf{x})$ is a vector field denoted by grad ϕ (or $\boldsymbol{\nabla}\phi$). Its i^{th}-component in the orthonormal basis is

$$(\text{grad } \phi)_i = \phi_{,i}, \tag{5.2}$$

so that

$$\text{grad } \phi = \phi_{,i}\mathbf{e}_i.$$

The gradient of a vector field $\mathbf{v}(\mathbf{x})$ is a 2-tensor field denoted by grad $\mathbf{v}$ (or $\boldsymbol{\nabla}\mathbf{v}$). Its ij^{th}-component in the orthonormal basis is

$$(\text{grad } \mathbf{v})_{ij} = v_{i,j}, \tag{5.3}$$

so that

$$\text{grad } \mathbf{v} = v_{i,j}\mathbf{e}_i \otimes \mathbf{e}_j.$$

The gradient of a scalar field ϕ in the particular direction of the unit vector $\mathbf{n}$ is denoted by $\partial\phi/\partial n$ and defined by

$$\frac{\partial \phi}{\partial n} = \boldsymbol{\nabla}\phi \cdot \mathbf{n}. \tag{5.4}$$

The **divergence** of a vector field $\mathbf{v}(\mathbf{x})$ is a scalar field denoted by div $\mathbf{v}$ (or $\boldsymbol{\nabla} \cdot \mathbf{v}$). It is given by

$$\text{div } \mathbf{v} = v_{i,i}. \tag{5.5}$$

The divergence of a 2-tensor field $\mathbf{A}(\mathbf{x})$ is a vector field denoted by div $\mathbf{A}$ (or $\boldsymbol{\nabla} \cdot \mathbf{A}$). Its i^{th}-component in the orthonormal basis is

$$(\text{div } \mathbf{A})_i = A_{ij,j} \tag{5.6}$$

so that

$$\text{div } \mathbf{A} = A_{ij,j}\mathbf{e}_i.$$

The **curl** of a vector field $\mathbf{v}(\mathbf{x})$ is a vector field denoted by curl $\mathbf{v}$ (or $\boldsymbol{\nabla} \times \mathbf{v}$). Its i^{th}-component in the orthonormal basis is

$$(\text{curl } \mathbf{v})_i = e_{ijk}v_{k,j} \tag{5.7}$$

so that

$$\text{curl } \mathbf{v} = e_{ijk}v_{k,j}\mathbf{e}_i.$$

The **Laplacian**s of a scalar field $\phi(\mathbf{x})$, a vector field $\mathbf{v}(\mathbf{x})$ and a 2-tensor field $\mathbf{A}(\mathbf{x})$ are the scalar, vector and 2-tensor fields with components

$$\nabla^2\phi = \phi_{,kk}, \qquad (\nabla^2\mathbf{v})_i = v_{i,kk}, \qquad (\nabla^2\mathbf{A})_{ij} = A_{ij,kk}, \tag{5.8}$$

5.2 Integral theorems

Let $\mathcal{D}$ be an arbitrary regular sub-region of the region $\mathcal{R}$. The **divergence theorem** allows one to relate a surface integral on $\partial\mathcal{D}$ to a volume integral on $\mathcal{D}$. In particular, for a scalar field $\phi(\mathbf{x})$

$$\int_{\partial D} \phi\mathbf{n}\, dA = \int_D \boldsymbol{\nabla}\phi\, dV \qquad \text{or} \qquad \int_{\partial D} \phi n_k\, dA = \int_D \phi_{,k}\, dV. \tag{5.9}$$

Likewise for a vector field $\mathbf{v}(\mathbf{x})$ one has

$$\int_{\partial D} \mathbf{v}\cdot\mathbf{n}\, dA = \int_D \boldsymbol{\nabla}\cdot\mathbf{v}\, dV \qquad \text{or} \qquad \int_{\partial D} v_k n_k\, dA = \int_D v_{k,k}\, dV, \tag{5.10}$$

as well as

$$\int_{\partial D} \mathbf{v}\otimes\mathbf{n}\, dA = \int_D \boldsymbol{\nabla}\mathbf{v}\, dV \qquad \text{or} \qquad \int_{\partial D} v_i n_k\, dA = \int_D v_{i,k}\, dV. \tag{5.11}$$

More generally for a n-tensor field $\mathbb{T}(\mathbf{x})$ the divergence theorem gives

$$\int_{\partial D} \mathbb{T}_{i_1 i_2 \ldots i_n}\, n_k\, dA = \int_D \frac{\partial}{\partial x_k}(\mathbb{T}_{i_1 i_2 \ldots i_n})\, dV \tag{5.12}$$

where some of the subscripts $i_1, i_2, \ldots, i_n$ may be repeated and one of them might equal k.

5.3 Localization

Certain physical principles are described to us in terms of equations that hold on an arbitrary portion of a body, i.e. in terms of an integral over a subregion $\mathcal{D}$ of $\mathcal{R}$. It is often useful to derive an equivalent statement of such a principle in terms of equations that must hold at each point $\mathbf{x}$ in the body. In what follows, we shall frequently have need to do this, i.e. convert a "global principle" to an equivalent "local field equation".

Consider for example the scalar field $\phi(\mathbf{x})$ that is defined and continuous at all $\mathbf{x} \in \mathcal{R} + \partial\, \mathcal{R}$ and suppose that

$$\int_{\mathcal{D}} \phi(\mathbf{x})\ \mathrm{d}V = 0 \qquad \text{for } \textit{all} \text{ subregions } \mathcal{D} \subset \mathcal{R}. \tag{5.13}$$

We will show that this "global principle" is equivalent to the "local field equation"

$$\phi(\mathbf{x}) = 0 \qquad \text{at every point } \mathbf{x} \in \mathbf{R}. \tag{5.14}$$

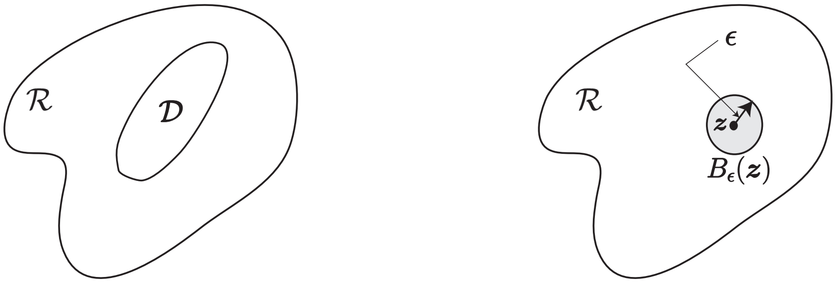


Figure 5.1: The region $\mathcal{R}$, a subregion $\mathcal{D}$ and a neighborhood $B_\epsilon(\mathbf{z})$ of the point $\mathbf{z}$.

We will prove this by contradiction. Suppose that (5.14) does *not* hold. This implies that there is a point, say $\mathbf{z} \in \mathcal{R}$, at which $\phi(\mathbf{z}) \neq 0$. Suppose that ϕ is positive at this point: $\phi(\mathbf{z}) > 0$. Since we are told that ϕ is continuous, ϕ is necessarily (strictly) positive in some neighborhood of $\mathbf{z}$ as well. Let $B_\epsilon(\mathbf{z})$ be a sphere with its center at $\mathbf{z}$ and radius $\epsilon > 0$. We can always choose ϵ sufficiently small so that $B_\epsilon(\mathbf{z})$ is a sufficiently small neighborhood of $\mathbf{z}$ and

$$\phi(\mathbf{x}) > 0 \qquad \text{at all } \mathbf{x} \in B_\epsilon(\mathbf{z}). \tag{5.15}$$

Now pick a region $\mathcal{D}$ which is a subset of $B_\epsilon(\mathbf{z})$. Then $\phi(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{D}$. Integrating ϕ over this $\mathcal{D}$ gives

$$\int_{\mathcal{D}} \phi(\mathbf{x})\ \mathrm{d}V > 0 \tag{5.16}$$

thus contradicting (5.13). An entirely analogous calculation can be carried out in the case $\phi(\mathbf{z}) < 0$. Thus our starting assumption must be false and (5.14) must hold.

5.4 Worked Examples.

In all of the examples below the region $\mathcal{R}$ will be a bounded regular region and its boundary $\partial\mathcal{R}$ will be smooth. All fields are defined on this region and are as smooth as in necessary.

In some of the examples below, we are asked to establish certain results for vector and tensor fields. When it is more convenient, we will carry out our calculations by first picking and fixing a basis, and then working with the components in that basis. If necessary, we will revert back to the vector and tensor fields at the end. We shall do this frequently in what follows and will not bother to explain this strategy each time.

Example 5.1: Calculate the gradient of the scalar-valued function $\phi(\mathbf{x}) = \mathbf{A}\mathbf{x}\cdot\mathbf{x}$ where $\mathbf{A}$ is a constant 2-tensor.

Solution: Writing ϕ in terms of components

$$\phi = A_{ij}x_ix_j.$$

Calculating the partial derivative of ϕ with respect to x_k yields

$$\phi_{,k} = A_{ij}(x_ix_j)_{,k} = A_{ij}(x_{i,k}x_j + x_ix_{j,k}) = A_{ij}(\delta_{ik}x_j + x_i\delta_{jk}) = A_{kj}x_j + A_{ik}x_i = (A_{kj} + A_{jk})x_j$$

or equivalently $\boldsymbol{\nabla}\phi = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$.

Example 5.2: Let $\mathbf{v}(\mathbf{x})$ be a vector field and let $v_i(x_1, x_2, x_3)$ be the i^{th}-component of $\mathbf{v}$ in a fixed orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. For each i and j define

$$F_{ij} = v_{i,j}.$$

Show that F_{ij} are the components of a 2-tensor.

Solution: Since $\mathbf{v}$ and $\mathbf{x}$ are 1-tensors, their components obey the transformation rules

$$v_i' = Q_{ik}v_k,\ v_i = Q_{ki}v_k' \qquad \text{and} \qquad x_j' = Q_{jk}x_k,\ x_\ell = Q_{j\ell}x_j'$$

Therefore

$$F_{ij}' = \frac{\partial v_i'}{\partial x_j'} = \frac{\partial v_i'}{\partial x_\ell}\frac{\partial x_\ell}{\partial x_j'} = \frac{\partial v_i'}{\partial x_\ell}Q_{j\ell} = \frac{\partial(Q_{ik}v_k)}{\partial x_\ell}Q_{j\ell} = Q_{ik}Q_{j\ell}\frac{\partial v_k}{\partial x_\ell} = Q_{ik}Q_{j\ell}F_{k\ell},$$

which is the transformation rule for a 2-tensor.

Example 5.3: If $\phi(\mathbf{x})$, $\mathbf{u}(\mathbf{x})$ and $\mathbf{A}(\mathbf{x})$ are a scalar, vector and 2-tensor fields respectively. Establish the identities

a. div $(\phi\mathbf{u}) = \mathbf{u}\cdot$ grad $\phi + \phi$ div $\mathbf{u}$

b. grad $(\phi \mathbf{u}) = \mathbf{u} \otimes \text{grad}\ \phi + \phi\ \text{grad}\ \mathbf{u}$

c. div $(\phi \mathbf{A}) = \mathbf{A}\ \text{grad}\ \phi + \phi\ \text{div}\ \mathbf{A}$

Solution:

a. In terms of components we are asked to show that $(\phi u_i)_{,i} = u_i \phi_{,i} + \phi\ u_{i,i}$. This follows immediately by expanding $(\phi u_i)_{,i}$ using the chain rule.

b. In terms of components we are asked to show that $(\phi u_i)_{,j} = u_i \phi_{,j} + \phi\ u_{i,j}$. Again, this follows immediately by expanding $(\phi u_i)_{,j}$ using the chain rule.

c. In terms of components we are asked to show that $(\phi A_{ij})_{,j} = A_{ij} \phi_{,j} + \phi\ A_{ij,j}$. Again, this follows immediately by expanding $(\phi A_{ij})_{,j}$ using the chain rule.

Example 5.4: If $\phi(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ are a scalar and vector field respectively, show that

$$\boldsymbol{\nabla} \times (\phi \mathbf{v}) = \phi(\boldsymbol{\nabla} \times \mathbf{v}) - \mathbf{v} \times \boldsymbol{\nabla}\phi \tag{i}$$

Solution: Recall that the curl of a vector field $\mathbf{u}$ can be expressed as $\boldsymbol{\nabla} \times \mathbf{u} = e_{ijk} u_{k,j} \mathbf{e}_i$ where $\mathbf{e}_i$ is a fixed basis vector. Thus evaluating $\boldsymbol{\nabla} \times (\phi \mathbf{v})$:

$$\boldsymbol{\nabla} \times (\phi \mathbf{v}) \ = \ e_{ijk}\ (\phi v_k)_{,j}\ \mathbf{e}_i \ = \ e_{ijk}\ \phi\ v_{k,j}\ \mathbf{e}_i + e_{ijk}\ \phi_{,j}\ v_k\ \mathbf{e}_i = \phi\ \boldsymbol{\nabla} \times \mathbf{v} + \boldsymbol{\nabla}\phi \times \mathbf{v} \tag{ii}$$

from which the desired result follows because $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$.

Example 5.5: Let $\mathbf{u}(\mathbf{x})$ be a vector field and define a second vector field $\boldsymbol{\xi}(\mathbf{x})$ by $\boldsymbol{\xi}(\mathbf{x}) = \text{curl}\ \mathbf{u}(\mathbf{x})$. Show that

a. $\boldsymbol{\nabla} \cdot \boldsymbol{\xi} = 0$;

b. $(\boldsymbol{\nabla}\mathbf{u} - \boldsymbol{\nabla}\mathbf{u}^T)\mathbf{a} = \boldsymbol{\xi} \times \mathbf{a}$ for any vector field $\mathbf{a}(\mathbf{x})$; and

c. $\boldsymbol{\xi} \cdot \boldsymbol{\xi} = \boldsymbol{\nabla}\mathbf{u} \cdot \boldsymbol{\nabla}\mathbf{u} - \boldsymbol{\nabla}\mathbf{u} \cdot \boldsymbol{\nabla}\mathbf{u}^T$

Solution: Recall that in terms of its components, $\boldsymbol{\xi} = \text{curl}\ \mathbf{u} = \boldsymbol{\nabla} \times \mathbf{u}$ can be expressed as

$$\xi_i = e_{ijk}\ u_{k,j}\ . \tag{i}$$

a. A direct calculation gives

$$\boldsymbol{\nabla} \cdot \boldsymbol{\xi} = \xi_{i,i} = (e_{ijk}\ u_{k,j})_{,i} = e_{ijk}\ u_{k,ji} = 0 \tag{ii}$$

where in the last step we have used the fact that e_{ijk} is skew-symmetric in the subscripts i, j, and $u_{k,ji}$ is symmetric in the subscripts i, j (since the order of partial differentiation can be switched) and therefore their product vanishes.

b. Multiplying both sides of (i) by e_{ipq} gives

$$e_{ipq}\,\xi_i = e_{ipq}\,e_{ijk}\,u_{k,j} = (\delta_{pj}\,\delta_{qk} - \delta_{pk}\,\delta_{qj})\,u_{k,j} = u_{q,p} - u_{p,q}, \tag{iii}$$

where we have made use of the identity $e_{ipq}\,e_{ijk} = \delta_{pj}\,\delta_{qk} - \delta_{pk}\,\delta_{qj}$ between the alternator and the Kronecker delta infroduced in (1.49) as well as the substitution rule. Multiplying both sides of this by a_q and using the fact that $e_{ipq} = -e_{piq}$ gives

$$e_{piq}\,\xi_i a_q = (u_{p,q} - u_{q,p})a_q, \tag{iv}$$

or $\boldsymbol{\xi} \times \mathbf{a} = (\boldsymbol{\nabla}\mathbf{u} - \boldsymbol{\nabla}\mathbf{u}^T)\mathbf{a}$.

c. Since $(\boldsymbol{\nabla}\mathbf{u})_{ij} = u_{i,j}$ and the inner product of two 2-tensors is $\mathbf{A}\cdot\mathbf{B} = A_{ij}B_{ij}$, the right-hand side of the equation we are asked to establish can be written as $\boldsymbol{\nabla}\mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{u} - \boldsymbol{\nabla}\mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{u}^T = (\boldsymbol{\nabla}\mathbf{u})_{ij}(\boldsymbol{\nabla}\mathbf{u})_{ij} - (\boldsymbol{\nabla}\mathbf{u})_{ij}(\boldsymbol{\nabla}\mathbf{u})_{ji} = u_{i,j}u_{i,j} - u_{i,j}u_{j,i}$. The left-hand side on the hand is $\boldsymbol{\xi}\cdot\boldsymbol{\xi} = \xi_i\,\xi_i$.

Using (i), the aforementioned identity between the alternator and the Kronecker delta, and the substitution rule leads to the desired result as follows:

$$\xi_i\,\xi_i = (e_{ijk}\,u_{k,j})\,(e_{ipq}\,u_{p,q}) = (\delta_{jp}\,\delta_{kq} - \delta_{jq}\,\delta_{kp})\,u_{k,j}\,u_{p,q} = u_{q,p}\,u_{p,q} - u_{p,q}\,u_{p,q}\ . \tag{v}$$

Example 5.6: Let $\mathbf{u}(\mathbf{x}), \mathbf{E}(\mathbf{x})$ and $\mathbf{S}(\mathbf{x})$ be, respectively, a vector and two 2-tensor fields. These fields are related by

$$\mathbf{E} = \frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\right), \qquad \mathbf{S} = 2\mu\mathbf{E} + \lambda\ \text{trace}(\mathbf{E})\ \mathbf{1}, \tag{i}$$

where λ and μ are constants. Suppose that

$$\mathbf{u}(\mathbf{x}) = b\,\frac{\mathbf{x}}{r^3} \qquad \text{where} \quad r = |\mathbf{x}|,\ |\mathbf{x}| \neq 0, \tag{ii}$$

and b is a constant. Use $(\text{i})_1$ to calculate the field $\mathbf{E}(\mathbf{x})$ corresponding to the field $\mathbf{u}(\mathbf{x})$ given in (ii), and then use $(\text{i})_2$ to calculate the associated field $\mathbf{S}(\mathbf{x})$. Thus verify that the field $\mathbf{S}(\mathbf{x})$ corresponding to (ii) satisfies the differential equation:

$$\text{div}\ \mathbf{S} = \mathbf{o}, \qquad |\mathbf{x}| \neq 0. \tag{iii}$$

Solution: We proceed in the manner suggested in the problem statement by first using $(\text{i})_1$ to calculate the $\mathbf{E}$ corresponding to the $\mathbf{u}$ given by (ii); substituting the result into $(\text{i})_2$ gives the corresponding $\mathbf{S}$; and finally we can then check whether or not this $\mathbf{S}$ satisfies (iii).

In components,

$$E_{ij} = \frac{1}{2}\left(u_{i,j} - u_{j,i}\right), \tag{iv}$$

and therefore we begin by calculting $u_{i,j}$. For this, it is convenient to first calculate $\partial r/\partial x_j = r_{,j}$. Observe by differentiating $r^2 = |\mathbf{x}|^2 = x_i\,x_i$ that

$$2rr_{,j} = 2x_{i,j}\,x_i = 2\delta_{ij}x_i\ = 2x_j, \tag{v}$$

and therefore

$$r_{,j} = \frac{x_j}{r}. \tag{vi}$$

Now differentiating the given vector field $u_i = bx_i/r^3$ with respect to x_j gives

$$\begin{aligned} u_{i,j} &= \frac{b}{r^3} x_{i,j} + b x_i\,(r^{-3})_{,j} &&= \frac{b}{r^3}\,\delta_{ij} - 3b\frac{x_i}{r^4}\,r_{,j} \\ &= b\,\frac{\delta_{ij}}{r^3} - 3b\,\frac{x_i}{r^4}\,\frac{x_j}{r} &&= b\,\frac{\delta_{ij}}{r^3} - 3b\,\frac{x_i\,x_j}{r^5}. \end{aligned} \tag{vii}$$

Substituting this into (iv) gives us E_{ij}:

$$E_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) = b\left(\frac{\delta_{ij}}{r^3} - 3\,\frac{x_i\,x_j}{r^5}\right). \tag{viii}$$

Next, substituting (viii) into $(\text{i})_2$, gives us S_{ij}:

$$\begin{aligned} S_{ij} &= 2\mu\,E_{ij} + \lambda E_{kk}\delta_{ij} \;=\; 2\mu b\left(\frac{\delta_{ij}}{r^3} - \frac{3x_i\,x_j}{r^5}\right) + \lambda b\left(\frac{\delta_{kk}}{r^3} - 3\,\frac{x_k\,x_k}{r^5}\right)\delta_{ij} \\ &= 2\mu b\left(\frac{\delta_{ij}}{r^3} - \frac{3x_i\,x_j}{r^5}\right) + \lambda b\left(\frac{3}{r^3} - 3\,\frac{r^2}{r^5}\right)\delta_{ij} \;=\; 2\mu b\left(\frac{\delta_{ij}}{r^3} - \frac{3x_i\,x_j}{r^5}\right). \end{aligned} \tag{ix}$$

Finally we use this to calculate $\partial S_{ij}/\partial x_j = S_{ij,j}$:

$$\begin{aligned} \frac{1}{2\mu b}\,S_{ij,j} &= \delta_{ij}\,(r^{-3})_{,j} - \frac{3}{r^5}\,(x_i\,x_j)_{,j} - 3x_i\,x_j\,(r^{-5})_{,j} \\ &= \delta_{ij}\,\left(-\frac{3}{r^4}r_{,j}\right) - \frac{3}{r^5}\,(\delta_{ij}\,x_j + x_i\,\delta_{jj}) - 3x_i\,x_j\left(-\frac{5}{r^6}\,r_{,j}\right) \\ &= -3\,\frac{\delta_{ij}}{r^4}\,\frac{x_j}{r} - \frac{3}{r^5}\,(x_i + 3x_i) + \frac{15x_i\,x_j}{r^6}\,\frac{x_j}{r} \\ &= 0. \end{aligned} \tag{x}$$

Example 5.7: Show that

$$\int_{\partial R} \mathbf{x}\otimes\mathbf{n}\,dA = V\mathbf{I}, \tag{i}$$

where V is the volume of the region R, and $\mathbf{x}$ is the position vector of a typical point in $R + \partial R$.

Solution: In terms of components in a fixed basis, we have to show that

$$\int_{\partial R} x_i n_j\;dA = V\delta_{ij}. \tag{ii}$$

The result follows immediately by using the divergence theorem (5.11):

$$\int_{\partial R} x_i n_j\;dA = \int_R x_{i,j}\;dV = \int_R \delta_{ij}\;dV \quad = \delta_{ij}\int_R dV \quad = \delta_{ij}V. \tag{iii}$$

Example 5.8: Let $\mathbf{A}(\mathbf{x})$ be a 2-tensor field with the property that

$$\int_{\partial\mathcal{D}} \mathbf{A}(\mathbf{x})\mathbf{n}(\mathbf{x})\;\mathrm{d}A \;=\; \mathbf{o} \qquad \text{for all subregions } \mathcal{D}\subset\;\mathcal{R}, \tag{i}$$

where $\mathbf{n}(\mathbf{x})$ is the unit outward normal vector at a point $\mathbf{x}$ on the boundary $\partial\mathcal{D}$. Show that (i) holds if and only if div $\mathbf{A} = \mathbf{o}$ at each point $\mathbf{x} \in \mathcal{R}$.

<u>*Solution:*</u> In terms of components in a fixed basis, we are told that

$$\int_{\partial\mathcal{D}} A_{ij}(\mathbf{x}) n_j(\mathbf{x}) \ \mathrm{d}A \ = \ 0 \qquad \text{for all subregions } \mathcal{D} \subset \mathcal{R}. \tag{ii}$$

By using the divergence theorem (5.12), this implies that

$$\int_{\mathcal{D}} A_{ij,j} \ \mathrm{d}V \ = \ 0 \qquad \text{for all subregions } \mathcal{D} \subset \mathcal{R}. \tag{iii}$$

If $A_{ij,j}$ is continuous on $\mathcal{R}$, the result established in the previous problem allows us to conclude that

$$A_{ij,j} \ = \ 0 \qquad \text{at each } \mathbf{x} \in \mathcal{R}. \tag{iv}$$

Conversely if (iv) holds, one can easily reverse the preceding steps to conclude that then (i) also holds. This shows that (iv) is both necessary and sufficient for (i) to hold.

<u>*Example 5.9:*</u> Let $\mathbf{A}(\mathbf{x})$ be a 2-tensor field which satisfies the differential equation div $\mathbf{A} = \mathbf{o}$ at each point in $\mathcal{R}$. Suppose that in addition

$$\int_{\partial\mathcal{D}} \mathbf{x} \times \mathbf{A}\mathbf{n} \ \mathrm{dA} \ = \mathbf{o} \quad \text{for all subregions } \mathcal{D} \subset \mathcal{R}.$$

Show that $\mathbf{A}$ must be a symmetric 2-tensor.

<u>*Solution:*</u> In terms of components we are given that

$$\int_{\partial\mathcal{D}} e_{ijk} x_j A_{kp} n_p \ \mathrm{dA} \ = 0,$$

which on using the divergence theorem yields

$$\int_{\mathcal{D}} e_{ijk} (x_j A_{kp})_{,p} \ \mathrm{dV} \ = \int_{\mathcal{D}} e_{ijk} [\delta_{jp} A_{kp} + x_j A_{kp,p}] \ \mathrm{dV} \ = 0.$$

We are also given that $A_{ij,j} = 0$ at each point in $\mathcal{R}$ and so the preceding equation simplifies, after using the substitution rule, to

$$\int_{\mathcal{D}} e_{ijk} A_{kj} \ \mathrm{dV} \ = 0.$$

Since this holds for all subregions $\mathcal{D} \subset \mathcal{R}$ we can localize it to

$$e_{ijk} A_{kj} \ = 0 \quad \text{at each } \mathbf{x} \in \mathcal{R}.$$

Finally, multiplying both sides by e_{ipq} and using the identity $e_{ipq} e_{ijk} = \delta_{pj}\delta_{qk} - \delta_{pk}\delta_{qj}$ in (1.49) yields

$$(\delta_{pj}\delta_{qk} - \delta_{pk}\delta_{qj}) A_{kj} = A_{qp} - A_{pq} = 0$$

and so $\mathbf{A}$ is symmetric.

<u>*Example 5.10:*</u> Let $\varepsilon_1(x_1, x_2)$ and $\varepsilon_2(x_1, x_2)$ be defined on a simply connected two-dimensional domain $\mathcal{R}$. Find necessary and sufficient conditions under which there exists a function $u(x_1, x_2)$ such that

$$u_{,1} = \varepsilon_1, \qquad u_{,2} = \varepsilon_2 \quad \text{for all } (x_1, x_2) \in \mathcal{R}. \tag{i}$$

<u>*Solution*</u>: In the presence of sufficient smoothness, the order of partial differentiation does not matter and so we necessarily have $u_{,12} = u_{,21}$. Therefore a necessary condition for (i) to hold is that $\varepsilon_1, \varepsilon_2$ obey

$$\varepsilon_{1,2} = \varepsilon_{2,1} \quad \text{for all } (x_1, x_2) \in \mathcal{R}. \tag{ii}$$

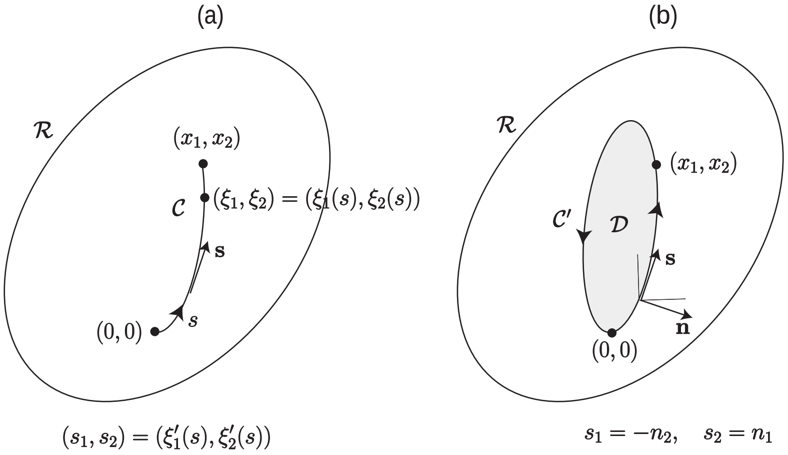


Figure 5.2: (a) Path $\mathcal{C}$ from $(0, 0)$ to (x_1, x_2). The curve is parameterized by arc length s as $\xi_1 = \xi_1(s), \xi_2 = \xi_2(s),\ 0 \leq s \leq s_0$. The unit tangent vector on $\mathcal{S}$, $\mathbf{s}$, has components (s_1, s_2). (b) A closed path $\mathcal{C}'$ passing through $(0, 0)$ and (x_1, x_2) and coinciding with $\mathcal{C}$ over part of its length. The unit outward normal vector on $\mathcal{S}$ is $\mathbf{n}$, and it has components (n_1, n_2)

.

To show that (ii) is also sufficient for the existence of u, we shall provide a formula for explicitly calculating the function u in terms of the given functions ε_1 and ε_2. Let $\mathcal{C}$ be an arbitrary regular oriented curve in $\mathcal{R}$ that connects $(0, 0)$ to (x_1, x_2). A generic point on the curve is denoted by (ξ_1, ξ_2) and the curve is characterized by the parameterization

$$\xi_1 = \xi_1(s), \xi_2 = \xi_2(s), \quad 0 \leq s \leq s_0, \tag{iii}$$

where s is arc length on $\mathcal{C}$ and $(\xi_1(0), \xi_2(0)) = (0, 0)$ and $(\xi_1(s_0), \xi_2(s_0)) = (x_1, x_2)$. We will show that the function

$$u(x_1, x_2) = \int_0^{s_0} \Big(\varepsilon_1(\xi_1(s), \xi_2(s))\xi_1'(s) + \varepsilon_2(\xi_1(s), \xi_2(s))\xi_2'(s) \Big) \mathrm{ds} \tag{iv}$$

satisfies the requirement (i) when (ii) holds.

To see this we must first show that the integral (iv) does in fact define a function of (x_1, x_2), i.e. that it does not depend on the path of integration. (Note that if a function u satisfies (i). then so does the function $u +$ *constant* and so the dependence on the arbitrary starting point of the integral is to be expected.) Thus consider a closed path $\mathcal{C}'$ that starts and ends at $(0, 0)$ and passes through (x_1, x_2) as sketched in Figure NNN (b). We need to show that

$$\int_{\mathcal{C}'} \Big(\varepsilon_1(\xi_1(s), \xi_2(s))\xi_1'(s) + \varepsilon_2(\xi_1(s), \xi_2(s))\xi_2'(s)\Big)\mathrm{ds} = 0. \tag{v}$$

Recall that $(\xi_1'(s), \xi_2'(s))$ are the components of the unit tangent vector on $\mathcal{C}'$ at the point $(\xi_1(s), \xi_2(s))$: $s_1 = \xi_1'(s), s_2 = \xi_2'(s)$. Observe further from the figure that the components of the unit tangent vector $\mathbf{s}$ and the unit outward normal vector $\mathbf{n}$ are related by $s_1 = -n_2$ and $s_2 = n_1$. Thus the left-hand side of (v) can be written as

$$\int_{\mathcal{C}'} \Big(\varepsilon_1 s_1 + \varepsilon_2 s_2\Big)\mathrm{ds} = \int_{\mathcal{C}'} \Big(\varepsilon_2 n_1 - \varepsilon_1 n_2\Big)\mathrm{ds} = \int_{\mathcal{D}'} \Big(\varepsilon_{2,1} - \varepsilon_{1,2}\Big)\mathrm{dA} \tag{vi}$$

where we have used the divergence theorem in the last step and $\mathcal{D}'$ is the region enclosed by $\mathcal{C}'$. In view of (ii), this last integral vanishes. Thus the integral (v) vanishes on any closed path $\mathcal{C}'$ and so the integral (iv) is independent of path and depends only on the end points. Thus (iv) does in fact define a function $u(x_1, x_2)$.

Finally it remains to show that the function (iv) satisfies the requirements (i). This is readily seen by writing (iv) in the form

$$u(x_1, x_2) = \int_{(0,0)}^{(x_1,x_2)} \Big(\varepsilon_1(\xi_1, \xi_2)\mathrm{d}\xi_1 + \varepsilon_2(\xi_1, \xi_2)\mathrm{d}\xi_2\Big) \tag{vii}$$

and then differentiating this with respect to x_1 and x_2.

Example 5.11: Let $a_1(x_1, x_2)$ and $a_2(x_1, x_2)$ be defined on a simply connected two-dimensional domain $\mathcal{R}$. Suppose that a_1 and a_2 satisfy the partial differential equation

$$a_{1,1}(x_1, x_2) + a_{2,2}(x_1, x_2) = 0 \quad \text{for all } (x_1, x_2) \in \mathcal{R}. \tag{i}$$

Show that (i) holds if and only if there is a function $\phi(x_1, x_2)$ such that

$$a_1(x_1, x_2) = \phi_{,2}(x_1, x_2), \quad a_2(x_1, x_2) = -\phi_{,1}(x_1, x_2). \tag{ii}$$

Solution: This is simply a restatement of the previous example in a form that will find useful in what follows.

Example 5.12: Find the most general vector field $\mathbf{u}(\mathbf{x})$ which satisfies the differential equation

$$\frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\right) = \boldsymbol{O} \quad \text{at all } \mathbf{x} \in \mathcal{R}. \tag{i}$$

Solution: In terms of components, $\boldsymbol{\nabla}\mathbf{u} = -\boldsymbol{\nabla}\mathbf{u}^T$ reads:

$$u_{i,j} = -u_{j,i}. \tag{ii}$$

Differentiating this with respect to x_k, and then changing the order of differentiation gives

$$u_{i,jk} = -u_{j,ik} = -u_{j,ki}.$$

However by (ii), $u_{j,k} = -u_{k,j}$. Using this and then changing the order of differentiation leads to

$$u_{i,jk} = -u_{j,ki} = u_{k,ji} = u_{k,ij}.$$

Again, by (ii), $u_{k,i} = -u_{i,k}$. Using this and changing the order of differentiation once again leads to

$$u_{i,jk} = u_{k,ij} = -u_{i,kj} = -u_{i,jk}.$$

It therefore follows that

$$u_{i,jk} = 0.$$

Integrating this once gives

$$u_{i,j} = C_{ij} \tag{iii}$$

where the C_{ij}'s are constants. Integrating this once more gives

$$u_i = C_{ij}x_j + c_i, \tag{iv}$$

where the c_i's are constants. The vector field $\mathbf{u}(\mathbf{x})$ must necessarily have this form if (ii) is to hold.

To examine sufficiency, substituting (iv) into (ii) shows that $[C]$ must be skew-symmetric. Thus in summary the most general vector field $\mathbf{u}(\mathbf{x})$ that satisfies (i) is

$$\mathbf{u}(\mathbf{x}) = \mathbf{C}\mathbf{x} + \mathbf{c}$$

where $\mathbf{C}$ is a constant skew-symmetric 2-tensor and $\mathbf{c}$ is a constant vector.

Example 5.13: Suppose that a scalar-valued function $f(\mathbf{A})$ is defined for all *symmetric* tensors $\mathbf{A}$. In terms of components in a fixed basis we have $f = f(A_{11}, A_{12}, A_{13}, A_{21}, \ldots A_{33})$. The partial derivatives of f with respect to A_{ij},

$$\frac{\partial f}{\partial A_{ij}} \ , \tag{i}$$

are the components of a 2-tensor. Is this tensor symmetric?

Solution: Consider, *for example*, the particular function $f = \mathbf{A} \cdot \mathbf{A} = A_{ij}A_{ij}$ which, when written out in components, reads:

$$f = f_1(A_{11}, A_{12}, A_{13}, A_{21}, \ldots A_{33}) = A_{11}^2 + A_{22}^2 + A_{33}^2 + 2A_{12}^2 + 2A_{23}^2 + 2A_{31}^2 \ . \tag{ii}$$

Proceeding formally and differentiating (ii) with respect to A_{12}, and separately with respect to A_{21}, gives

$$\frac{\partial f_1}{\partial A_{12}} = 4A_{12}, \qquad \frac{\partial f_1}{\partial A_{21}} = 0, \tag{iii}$$

which implies that $\partial f_1/\partial A_{12} \neq \partial f_1/\partial A_{21}$.

On the other hand, since A_{ij} is symmetric we can write

$$A_{ij} = \frac{1}{2}\left(A_{ij} + A_{ji}\right). \tag{iv}$$

Substituting (iv) into the formula (ii) for f gives $f = f_2(A_{11}, A_{12}, A_{13}, A_{21}, \ldots A_{33})$:

$$\begin{aligned} &f_2(A_{11}, A_{12}, A_{13}, A_{21}, \ldots A_{33}) \\ &= \quad A_{11}^2 + A_{22}^2 + A_{33}^2 + 2\left[\frac{1}{2}(A_{12} + A_{21})\right]^2 + 2\left[\frac{1}{2}(A_{23} + A_{31})\right]^2 + 2\left[\frac{1}{2}(A_{31} + A_{13})\right]^2 \ , \\ &= \quad A_{11}^2 + A_{22}^2 + A_{33}^2 + \frac{1}{2}A_{12}^2 + A_{12}A_{21} + \frac{1}{2}A_{21}^2 + \ldots + \frac{1}{2}A_{31}^2 + A_{31}A_{13} + \frac{1}{2}A_{13}^2 \ . \end{aligned} \tag{v}$$

Note that the values of $f_1[A] = f_2[A]$ for any symmetric matrix $[A]$. Differentiating f_2 leads to

$$\frac{\partial f_2}{\partial A_{12}} = A_{12} + A_{21}, \qquad \frac{\partial f_2}{\partial A_{21}} = A_{21} + A_{12} \ , \tag{vi}$$

and so now, $\partial f_2/\partial A_{12} = \partial f_2/\partial A_{21}$.

The source of the original difficulty is the fact that the 9 A_{ij}'s in the argument of f_1 are *not* independent variables since $A_{ij} = A_{ji}$; and yet we have been calculating partial derivatives as if they were independent. In fact, the original problem statement itself is ill-posed since we are asked to calculate $\partial f/\partial A_{ij}$ but told that $[A]$ is restricted to being symmetric.

Suppose that f_2 is defined by (v) for *all* matrices $[A]$ and not just symmetric matrices $[A]$. We see that the values of the functions f_1 and f_2 are equal at all symmetric matrices and so in going from $f_1 \to f_2$, we have effectively relaxed the constraint of symmetry and expanded the domain of definition of f to *all* matrices $[A]$. We may differentiate f_2 by treating the 9 A_{ij}'s to be independent and the result can then be evaluated at symmetric matrices. We assume that this is what was meant in the problem statement.

In general, if a function $f(A_{11}, A_{12}, \ldots A_{33})$ is expressed in symmetric form, by changing $A_{ij} \to \frac{1}{2}(A_{ij} + A_{ji})$, then $\partial f/\partial A_{ij}$ will be symmetric; but not otherwise. Throughout these volumes, whenever we encounter a function of a symmetric tensor, we shall always assume that it has been written in symmetric form; and therefore its derivative with respect to the tensor can be assumed to be symmetric.

Remark: We will encounter a similar situation involving tensors whose determinant is unity. On occasion we will have need to differentiate a function $g_1(\mathbf{A})$ defined for all tensors with $\det \mathbf{A} = 1$ and we shall do this by extending the definition of the given function and defining a second function $g_2(\mathbf{A})$ for all tensors; g_2 is defined such that $g_1(\mathbf{A}) = g_2(\mathbf{A})$ for all tensors with unit determinant. We then differentiate g_2 and evaluate the result at tensors with unit determinant.

References

1. P. Chadwick, *Continuum Mechanics*, Chapter 1, Sections 10 and 11, Dover, 1999.

2. M.E. Gurtin, *An Introduction to Continuum Mechanics*, Chapter 2, Academic Press, 1981.

3. L.A. Segel, *Mathematics Applied to Continuum Mechanics*, Section 2.3, Dover, New York, 1987.

Chapter 6

Orthogonal Curvilinear Coordinates

6.1 Introductory Remarks

The notes in this section are a somewhat simplified version of notes developed by Professor Eli Sternberg of Caltech. The discussion here, which is a general treatment of *orthogonal* curvilinear coordinates, is a compromise between a general tensorial treatment that includes oblique coordinate systems and an ad-hoc treatment of special orthogonal curvilinear coordinate systems. A summary of the main tensor analytic results of this section are given in equations (6.32) - (6.37) in terms of the scale factors h_i defined in (6.17) that relate the rectangular cartesian coordinates (x_1, x_2, x_3) to the orthogonal curvilinear coordinates $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$.

It is helpful to begin by reviewing a few aspects of the familiar case of *circular cylindrical coordinates*. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be a *fixed* orthonormal basis, and let O be a *fixed* point chosen as the origin. The point O and the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, together, constitute a *frame* which we denote by $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Consider a generic point P in $\mathbb{R}_3$ whose position vector relative to this origin O is $\mathbf{x}$. The *rectangular cartesian coordinates* of the point P in the frame $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are the components (x_1, x_2, x_3) of the position vector $\mathbf{x}$ in this basis.

We introduce circular cylindrical coordinates (r, θ, z) through the mappings

$$\left.\begin{array}{c} x_1 = r\cos\theta; \quad x_2 = r\sin\theta; \quad x_3 = z; \\ \text{for all } (r,\theta,z) \in [0,\infty) \times [0,2\pi) \times (-\infty,\infty). \end{array}\right\} \tag{6.1}$$

The mapping (6.1) is one-to-one except at $r = 0$ (i.e. $x_1 = x_2 = 0$). Indeed (6.1) may be explicitly inverted for $r > 0$ to give

$$r = \sqrt{x_1^2 + x_2^2}; \qquad \cos\theta = x_1/r,\ \sin\theta = x_2/r; \qquad z = x_3. \tag{6.2}$$

For a general set of orthogonal curvilinear coordinates one cannot, in general, explicitly invert the coordinate mapping in this way.

The Jacobian determinant of the mapping (6.1) is

$$\Delta(r,\theta,z) = \det \begin{bmatrix} \partial x_1/\partial r & \partial x_1/\partial\theta & \partial x_1/\partial z \\ \partial x_2/\partial r & \partial x_2/\partial\theta & \partial x_2/\partial z \\ \partial x_3/\partial r & \partial x_3/\partial\theta & \partial x_3/\partial z \end{bmatrix} = r \ \geq \ 0.$$

Note that $\Delta(r,\theta,z) = 0$ if and only if $r = 0$ and is otherwise strictly positive; this reflects the invertibility of (6.1) on $(r,\theta,z) \in (0,\infty) \times [0,2\pi) \times (-\infty,\infty)$, and the breakdown in invertibility at $r = 0$.

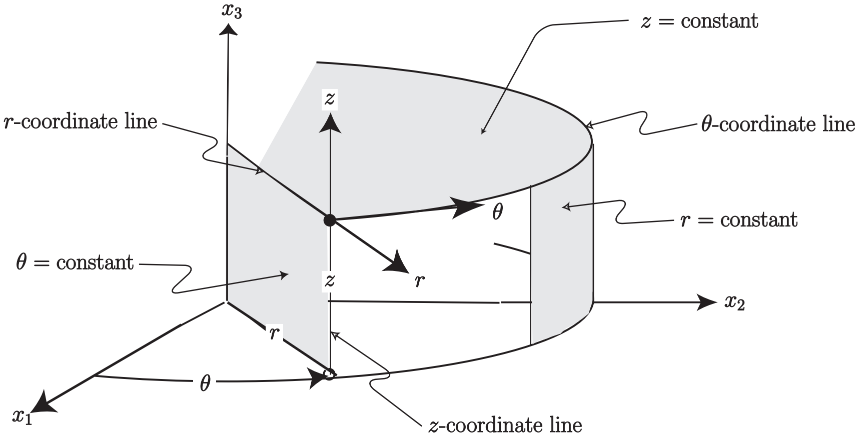


Figure 6.1: Circular cylindrical coordinates (r,θ,z).

The circular cylindrical coordinates (r,θ,z) admit the familiar geometric interpretation illustrated in Figure 6.1. In view of (6.2), one has:

$$\begin{array}{lllll} r & = & r_o & = & \text{constant}: \quad \text{circular cylinders, co} - \text{axial with } x_3 - \text{axis}, \\ \theta & = & \theta_o & = & \text{constant}: \quad \text{meridional half} - \text{planes through } x_3 - \text{axis}, \\ z & = & z_o & = & \text{constant}: \quad \text{planes perpendicular to } x_3 - \text{axis}. \end{array}$$

The above surfaces constitute a triply orthogonal family of *coordinate surfaces*; each "regular point" of $\mathbb{E}_3$ (i.e. a point at which $r > 0$) is the intersection of a unique triplet of (mutually perpendicular) coordinate surfaces. The *coordinate lines* are the pairwise intersections of the coordinate surfaces; thus for example as illustrated in Figure 6.1, the line along which a r-coordinate surface and a z-coordinate surface intersect is a θ-coordinate line. Along any coordinate line only one of the coordinates (r, θ, z) varies, while the other two remain constant.

In terms of the circular cylindrical coordinates the position vector $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{x}(r, \theta, z) = (r\cos\theta)\mathbf{e}_1 + (r\sin\theta)\mathbf{e}_2 + z\mathbf{e}_3. \tag{6.3}$$

The vectors

$$\partial\mathbf{x}/\partial r, \quad \partial\mathbf{x}/\partial\theta, \quad \partial\mathbf{x}/\partial z,$$

are tangent to the coordinate lines corresponding to r, θ and z respectively. The so-called metric coefficients h_r, h_θ, h_z denote the magnitudes of these vectors, i.e.

$$h_r = |\partial\mathbf{x}/\partial r|, \qquad h_\theta = |\partial\mathbf{x}/\partial\theta|, \qquad h_z = |\partial\mathbf{x}/\partial z|,$$

and so the unit tangent vectors corresponding to the respective coordinate lines r, θ and z are:

$$\mathbf{e}_r = \frac{1}{h_r}(\partial\mathbf{x}/\partial r), \qquad \mathbf{e}_\theta = \frac{1}{h_\theta}(\partial\mathbf{x}/\partial\theta), \qquad \mathbf{e}_z = \frac{1}{h_z}(\partial\mathbf{x}/\partial z).$$

In the present case one has $h_r = 1, h_\theta = r, h_z = 1$ and

$$\left.\begin{array}{rclcrcl}
\mathbf{e}_r & = & \dfrac{1}{h_r}(\partial\mathbf{x}/\partial r) & = & \cos\theta\ \mathbf{e}_1 & + & \sin\theta\ \mathbf{e}_2, \\
\mathbf{e}_\theta & = & \dfrac{1}{h_\theta}(\partial\mathbf{x}/\partial\theta) & = & -\sin\theta\ \mathbf{e}_1 & + & \cos\theta\mathbf{e}_2, \\
\mathbf{e}_z & = & \dfrac{1}{h_z}(\partial\mathbf{x}/\partial z) & = & \mathbf{e}_3. & &
\end{array}\right\}$$

The triplet of vectors $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z\}$ forms a *local* orthonormal basis at the point $\mathbf{x}$. They are local because they depend on the point $\mathbf{x}$; sometimes, when we need to emphasize this fact, we will write $\{\mathbf{e}_r(\mathbf{x}), \mathbf{e}_\theta(\mathbf{x}), \mathbf{e}_z(\mathbf{x})\}$.

In order to calculate the derivatives of various field quantities it is clear that we will need to calculate quantities such as $\partial\mathbf{e}_r/\partial r,\ \partial\mathbf{e}_r/\partial\theta, \ldots$ etc. ; and in order to calculate the *components* of these derivatives in the local basis we will need to calculate quantities of the form

$$\begin{array}{llll}
\mathbf{e}_r \cdot (\partial\mathbf{e}_r/\partial r), & \mathbf{e}_\theta \cdot (\partial\mathbf{e}_r/\partial r), & \mathbf{e}_z \cdot (\partial\mathbf{e}_r/\partial r), & \\
\mathbf{e}_r \cdot (\partial\mathbf{e}_\theta/\partial r), & \mathbf{e}_\theta \cdot (\partial\mathbf{e}_\theta/\partial r), & \mathbf{e}_z \cdot (\partial\mathbf{e}_\theta/\partial r), & \ldots\text{etc.}
\end{array} \tag{6.4}$$

Much of the analysis in the general case to follow, leading eventually to Equation (6.30) in Sub-section 6.2.4, is devoted to calculating these quantities.

Notation: As far as possible, we will consistently denote the fixed cartesian coordinate system and all components and quantities associated with it by symbols such as $x_i, \mathbf{e}_i, f(x_1, x_2, x_3)$, $v_i(x_1, x_2, x_3)$, $A_{ij}(x_1, x_2, x_3)$ etc. and we shall consistently denote the local curvilinear coordinate system and all components and quantities associated with it by similar symbols with "hats" over them, e.g. $\hat{x}_i, \hat{\mathbf{e}}_i, \hat{f}(\hat{x}_1, \hat{x}_2, \hat{x}_3), \hat{v}_i(\hat{x}_1, \hat{x}_2, \hat{x}_3), \widehat{A}_{ij}(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ etc.

6.2 General Orthogonal Curvilinear Coordinates

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be a *fixed* right-handed orthonormal basis, let O be the *fixed* point chosen as the origin and let $\{O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the associated frame. The rectangular cartesian coordinates of the point with position vector $\mathbf{x}$ in this frame are

$$(x_1, x_2, x_3) \quad \text{where } x_i = \mathbf{x} \cdot \mathbf{e}_i.$$

6.2.1 Coordinate transformation. Inverse transformation.

We introduce curvilinear coordinates $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ through a triplet of scalar mappings

$$x_i = x_i(\hat{x}_1, \hat{x}_2, \hat{x}_3) \quad \text{for all } (\hat{x}_1, \hat{x}_2, \hat{x}_3) \in \hat{R}, \tag{6.5}$$

where the domain of definition $\hat{R}$ is a subset of $\mathbb{E}_3$. Each curvilinear coordinate $\hat{x}_i$ belongs to some linear interval $\mathcal{L}_i$, and $\hat{R} = \mathcal{L}_1 \times \mathcal{L}_2 \times \mathcal{L}_3$. For example in the case of circular cylindrical coordinates we have $\mathcal{L}_1 = \{(\hat{x}_1 \mid 0 \leq \hat{x}_1 < \infty\}$, $\mathcal{L}_2 = \{(\hat{x}_2 \mid 0 \leq \hat{x}_2 < 2\pi\}$ and $\mathcal{L}_3 = \{(\hat{x}_3 \mid -\infty < \hat{x}_3 < \infty\}$, and the "box" $\hat{R}$ is given by $\hat{R} = \{(\hat{x}_1, \hat{x}_2, \hat{x}_3) \mid 0 \leq \hat{x}_1 < \infty, 0 \leq \hat{x}_2 < 2\pi, -\infty < \hat{x}_3 < \infty\}$. Observe that the "box" $\hat{R}$ includes some but possibly not all of its faces.

Equation (6.5) may be interpreted as a mapping of $\hat{R}$ into $\mathbb{E}_3$. We shall assume that (x_1, x_2, x_3) ranges over *all* of $\mathbb{E}_3$ as $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ takes on all values in $\hat{R}$. We assume further that the mapping (6.5) is one-to-one and sufficiently smooth in the interior of $\hat{R}$ so that the inverse mapping

$$\hat{x}_i = \hat{x}_i(x_1, x_2, x_3) \tag{6.6}$$

exists and is appropriately smooth at all (x_1, x_2, x_3) in the image of the interior of $\hat{R}$.

Note that the mapping (6.5) might not be uniquely invertible on some of the faces of $\hat{R}$ which are mapped into "singular" lines or surfaces in $\mathbb{E}_3$. (For example in the case of circular cylindrical coordinates, $\hat{x}_1 = r = 0$ is a singular surface; see Section 6.1.) Points that are not on a singular line or surface will be referred to as "regular points" of $\mathbb{E}_3$.

The *Jacobian* matrix $[J]$ of the mapping (6.5) has elements

$$J_{ij} = \frac{\partial x_i}{\partial \hat{x}_j} \tag{6.7}$$

and by the assumed smoothness and one-to-oneness of the mapping, the Jacobian determinant does not vanish on the interior of $\hat{R}$. Without loss of generality we can take therefore take it to be positive:

$$\det[J] = \frac{1}{6} e_{ijk} e_{pqr} \frac{\partial x_i}{\partial \hat{x}_p} \frac{\partial x_j}{\partial \hat{x}_q} \frac{\partial x_k}{\partial \hat{x}_r} \ > \ 0 \ . \tag{6.8}$$

The Jacobian matrix of the inverse mapping (6.6) is $[J]^{-1}$.

The *coordinate surface* $\hat{x}_i =$ constant is defined by

$$\hat{x}_i(x_1, x_2, x_3) = \hat{x}_i^o = \text{constant}, \quad i = 1, 2, 3;$$

the pairwise intersections of these surfaces are the corresponding *coordinate lines*, along which only one of the curvilinear coordinates varies. Thus every regular point of $\mathbb{E}_3$ is the point of intersection of a unique triplet of coordinate surfaces and coordinate lines, as is illustrated in Figure 6.2.

Recall that the tangent vector along an arbitrary regular curve

$$\Gamma : \mathbf{x} = \mathbf{x}(t), \quad (\alpha \leq t \leq \beta), \tag{6.9}$$

can be taken to be[1] $\dot{\mathbf{x}}(t) = \dot{x}_i(t)\mathbf{e}_i$; it is oriented in the direction of increasing t. Thus in the case of the special curve

$$\Gamma_1 : \mathbf{x} = \mathbf{x}(\hat{x}_1, c_2, c_3), \quad \hat{x}_1 \in \mathcal{L}_1, \quad c_2 = \text{constant}, \ c_3 = \text{constant},$$

corresponding to a $\hat{x}_1$-coordinate line, the tangent vector can be taken to be $\partial\mathbf{x}/\partial\hat{x}_1$. Generalizing this, $\partial\mathbf{x}/\partial\hat{x}_i$ are tangent vectors and

$$\hat{\mathbf{e}}_i = \frac{1}{|\partial\mathbf{x}/\partial\hat{x}_i|} \frac{\partial\mathbf{x}}{\partial\hat{x}_i} \quad \text{(no sum)} \tag{6.10}$$

[1] Here and in the sequel a superior dot indicates differentiation with respect to the parameter t.

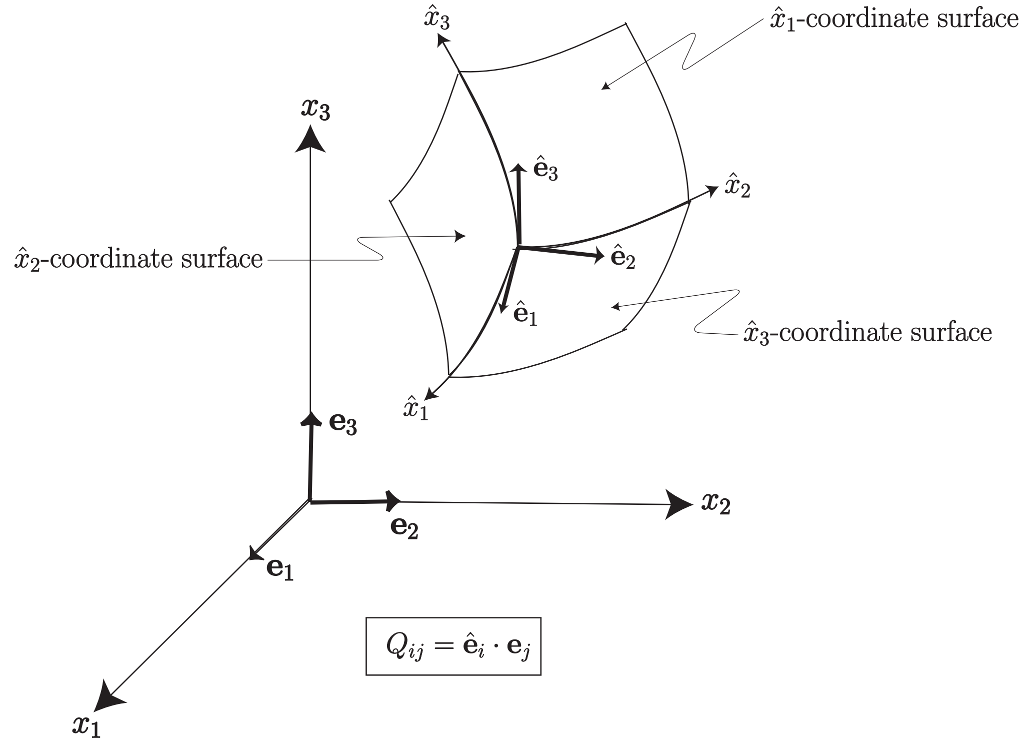


Figure 6.2: Orthogonal curvilinear coordinates $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ and the associated local orthonormal basis vectors $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$. Here $\hat{\mathbf{e}}_i$ is the unit tangent vector along the $\hat{x}_i$-coordinate line, the sense of $\hat{\mathbf{e}}_i$ being determined by the direction in which $\hat{x}_i$ increases. The proper orthogonal matrix $[Q]$ characterizes the rotational transformation relating this basis to the rectangular cartesian basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

are the *unit* tangent vectors along the $\hat{x}_i-$coordinate lines, both of which point in the sense of increasing $\hat{x}_i$. Since our discussion is limited to *orthogonal* curvilinear coordinate systems, we must require

$$\text{for } i \neq j: \quad \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = 0 \quad \text{or} \quad \frac{\partial \mathbf{x}}{\partial \hat{x}_i} \cdot \frac{\partial \mathbf{x}}{\partial \hat{x}_j} = 0 \qquad \text{or} \quad \frac{\partial x_k}{\partial \hat{x}_i} \cdot \frac{\partial x_k}{\partial \hat{x}_j} = 0. \tag{6.11}$$

6.2.2 Metric coefficients, scale moduli.

Consider again the arbitrary regular curve Γ parameterized by (6.9). If $s(t)$ is the arc-length of Γ, measured from an arbitrary fixed point on Γ, one has

$$|\dot{s}(t)| = |\dot{\mathbf{x}}(t)| = \sqrt{\dot{\mathbf{x}}(t) \cdot \dot{\mathbf{x}}(t)}. \tag{6.12}$$

One concludes from (6.12), (6.5) and the chain rule that

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\cdot\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \frac{\mathrm{d}x_k}{\mathrm{d}t}\cdot\frac{\mathrm{d}x_k}{\mathrm{d}t} = \left(\frac{\partial x_k}{\partial \hat{x}_i}\frac{\mathrm{d}\hat{x}_i}{\mathrm{d}t}\right)\cdot\left(\frac{\partial x_k}{\partial \hat{x}_j}\frac{\mathrm{d}\hat{x}_j}{\mathrm{d}t}\right) = \left(\frac{\partial x_k}{\partial \hat{x}_i}\cdot\frac{\partial x_k}{\partial \hat{x}_j}\right)\frac{\mathrm{d}\hat{x}_i}{\mathrm{d}t}\,\frac{\mathrm{d}\hat{x}_j}{\mathrm{d}t}.$$

where

$$\hat{x}_i(t) = \hat{x}_i(x_1(t), x_2(t), x_3(t)), \qquad (\alpha \le t \le \beta),$$

Thus

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = g_{ij}\,\frac{\mathrm{d}\hat{x}_i}{\mathrm{d}t}\frac{\mathrm{d}\hat{x}_j}{\mathrm{d}t} \qquad \text{or} \qquad (\mathrm{d}s)^2 = g_{ij}\,\mathrm{d}\hat{x}_i\,\mathrm{d}\hat{x}_j, \tag{6.13}$$

in which g_{ij} are the *metric coefficients* of the curvilinear coordinate system under consideration. They are defined by

$$g_{ij} = \frac{\partial \mathbf{x}}{\partial \hat{x}_i}\cdot\frac{\partial \mathbf{x}}{\partial \hat{x}_j} = \frac{\partial x_k}{\partial \hat{x}_i}\frac{\partial x_k}{\partial \hat{x}_j}. \tag{6.14}$$

Note that

$$g_{ij} = 0, \qquad (i \ne j), \tag{6.15}$$

as a consequence of the orthogonality condition (6.11). Observe that in terms of the Jacobian matrix $[J]$ defined earlier in (6.7) we can write $g_{ij} = J_{ki}J_{kj}$ or equivalently $[g] = [J]^T[J]$.

Because of (6.14) and (6.15) the metric coefficients can be written as

$$g_{ij} = h_{\underline{i}}h_{\underline{j}}\delta_{ij}, \tag{6.16}$$

where the *scale moduli* h_i are defined by[23]

$$h_i = \sqrt{g_{i\underline{i}}} = \sqrt{\frac{\partial x_k}{\partial \hat{x}_i}\frac{\partial x_k}{\partial \hat{x}_{\underline{i}}}} = \sqrt{\left(\frac{\partial x_1}{\partial \hat{x}_i}\right)^2 + \left(\frac{\partial x_2}{\partial \hat{x}_i}\right)^2 + \left(\frac{\partial x_3}{\partial \hat{x}_i}\right)^2} > 0, \tag{6.17}$$

noting that $h_i = 0$ is precluded by (6.8). The matrix of metric coefficients is therefore

$$[g] = \begin{pmatrix} h_1^2 & 0 & 0 \\ 0 & h_2^2 & 0 \\ 0 & 0 & h_3^2 \end{pmatrix}. \tag{6.18}$$

From (6.13), (6.14), (6.17) follows

$$(ds)^2 = (h_1 d\hat{x}_1)^2 + (h_2 d\hat{x}_2)^2 + (h_3 d\hat{x}_3)^2 \qquad \text{along } \Gamma, \tag{6.19}$$

[2] Here and henceforth the underlining of one of two or more repeated indices indicates suspended summation with respect to this index.

[3] Some authors such as Love define h_i as $1/\sqrt{g_{i\underline{i}}}$ instead of as $\sqrt{g_{i\underline{i}}}$

which reveals the geometric significance of the scale moduli, i.e.

$$h_i = \frac{\mathrm{d}s}{\mathrm{d}\hat{x}_i} \quad \text{along the } \hat{x}_i - \text{coordinate lines.} \tag{6.20}$$

It follows from (6.17), (6.10) and (6.11) that the unit vector $\hat{\mathbf{e}}_i$ can be expressed as

$$\hat{\mathbf{e}}_i = \frac{1}{h_{\underline{i}}} \frac{\partial \mathbf{x}}{\partial \hat{x}_i}, \tag{6.21}$$

and therefore the proper orthogonal matrix $[Q]$ relating the two sets of basis vectors is given by

$$Q_{ij} = \hat{\mathbf{e}}_i \cdot \mathbf{e}_j = \frac{1}{h_{\underline{i}}} \frac{\partial x_j}{\partial \hat{x}_i} \tag{6.22}$$

6.2.3 Inverse partial derivatives

In view of (6.5), (6.6) one has the identity

$$x_i = x_i(\hat{x}_1(x_1, x_2, x_3), \hat{x}_2(x_1, x_2, x_3), \hat{x}_3(x_1, x_2, x_3)),$$

so that from the chain rule,

$$\frac{\partial x_i}{\partial \hat{x}_k} \frac{\partial \hat{x}_k}{\partial x_j} = \delta_{ij}.$$

Multiply this by $\partial x_i / \partial \hat{x}_m$, noting the implied contraction on the index i, and use (6.14), (6.16) to confirm that

$$\frac{\partial x_j}{\partial \hat{x}_m} = g_{km} \frac{\partial \hat{x}_k}{\partial x_j} = h^2_{\underline{m}} \frac{\partial \hat{x}_m}{\partial x_j}.$$

Thus the inverse partial derivatives are given by

$$\frac{\partial \hat{x}_i}{\partial x_j} = \frac{1}{h^2_{\underline{i}}} \frac{\partial x_j}{\partial \hat{x}_i}. \tag{6.23}$$

By (6.22), the elements of the matrix $[Q]$ that relates the two coordinate systems can be written in the alternative form

$$Q_{ij} = \hat{\mathbf{e}}_i \cdot \mathbf{e}_j = h_{\underline{i}} \frac{\partial \hat{x}_i}{\partial x_j}. \tag{6.24}$$

Moreover, (6.23) and (6.17) yield the following alternative expressions for h_i:

$$h_i = 1 \Big/ \sqrt{\Big(\frac{\partial \hat{x}_i}{\partial x_1}\Big)^2 + \Big(\frac{\partial \hat{x}_i}{\partial x_2}\Big)^2 + \Big(\frac{\partial \hat{x}_i}{\partial x_3}\Big)^2}.$$

6.2.4 Components of $\partial\hat{\mathbf{e}}_i/\partial\hat{x}_j$ in the local basis $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$

In order to calculate the derivatives of various field quantities it is clear that we will need to calculate the quantities $\partial\hat{\mathbf{e}}_i/\partial\hat{x}_j$; and in order to calculate the *components* of these derivatives in the local basis we will need to calculate quantities of the form $\hat{\mathbf{e}}_k \cdot \partial\hat{\mathbf{e}}_i/\partial\hat{x}_j$. Calculating these quantities is an essential prerequisite for the transformation of basic tensor-analytic relations into arbitrary orthogonal curvilinear coordinates, and this subsection is devoted to this calculation.

From (6.21),

$$\frac{\partial\hat{\mathbf{e}}_i}{\partial\hat{x}_j} \cdot \hat{\mathbf{e}}_k = \left\{ -\frac{1}{h_{\underline{i}}^2}\frac{\partial h_i}{\partial\hat{x}_j}\frac{\partial\mathbf{x}}{\partial\hat{x}_i} + \frac{1}{h_{\underline{i}}}\frac{\partial^2\mathbf{x}}{\partial\hat{x}_i\partial\hat{x}_j} \right\} \cdot \frac{1}{h_{\underline{k}}}\frac{\partial\mathbf{x}}{\partial\hat{x}_k},$$

while by (6.14), (6.17),

$$\frac{\partial\mathbf{x}}{\partial\hat{x}_i} \cdot \frac{\partial\mathbf{x}}{\partial\hat{x}_j} = g_{ij} = \delta_{ij} h_{\underline{i}} h_{\underline{j}}. \qquad (6.25)$$

Therefore

$$\frac{\partial\hat{\mathbf{e}}_i}{\partial\hat{x}_j} \cdot \hat{\mathbf{e}}_k = -\frac{\delta_{ik}}{h_{\underline{i}}}\frac{\partial h_i}{\partial\hat{x}_j} + + \frac{1}{h_{\underline{i}} h_{\underline{k}}}\frac{\partial^2\mathbf{x}}{\partial\hat{x}_i\partial\hat{x}_k} \cdot \frac{\partial\mathbf{x}}{\partial x_k} \ . \qquad (6.26)$$

In order to express the second derivative term in (6.26) in terms of the scale-moduli and their first partial derivatives, we begin by differentiating (6.25) with respect to $\hat{x}_k$. Thus,

$$\frac{\partial^2\mathbf{x}}{\partial\hat{x}_i\partial\hat{x}_k} \cdot \frac{\partial\mathbf{x}}{\partial\hat{x}_j} + \frac{\partial^2\mathbf{x}}{\partial\hat{x}_j\partial\hat{x}_k} \cdot \frac{\partial\mathbf{x}}{\partial\hat{x}_i} = \delta_{ij}\frac{\partial}{\partial\hat{x}_k}(h_{\underline{i}} h_{\underline{j}}) \ . \qquad (6.27)$$

If we refer to (6.27) as (a), and let (b) and (c) be the identities resulting from (6.27) when (i, j, k) are replaced by (j, k, i) and (k, i, j), respectively, then $\frac{1}{2}\{$(b)+(c) -(a)$\}$ is readily found to yield

$$\frac{\partial^2\mathbf{x}}{\partial\hat{x}_i\partial\hat{x}_j} \cdot \frac{\partial\mathbf{x}}{\partial\hat{x}_k} = \frac{1}{2}\left\{ \delta_{\underline{jk}}\frac{\partial}{\partial\hat{x}_i}(h_j h_k) + \delta_{\underline{ki}}\frac{\partial}{\partial\hat{x}_j}(h_k h_i) - \delta_{\underline{ij}}\frac{\partial}{\partial\hat{x}_k}(h_i h_j) \right\}. \qquad (6.28)$$

Substituting (6.28) into (6.26) leads to

$$\frac{\partial\hat{\mathbf{e}}_i}{\partial\hat{x}_j} \cdot \hat{\mathbf{e}}_k = -\frac{\delta_{ik}}{h_{\underline{i}}}\frac{\partial h_i}{\partial\hat{x}_j} + \frac{1}{2h_{\underline{i}} h_{\underline{k}}}\left\{ \delta_{\underline{j}k}\frac{\partial}{\partial\hat{x}_i}(h_j h_k) + \delta_{ki}\frac{\partial}{\partial\hat{x}_j}(h_k h_i) - \delta_{i\underline{j}}\frac{\partial}{\partial\hat{x}_k}(h_i h_j) \right\}. \qquad (6.29)$$

Equation (6.29) provides the explicit expressions for the terms $\partial\hat{\mathbf{e}}_i/\partial\hat{x}_j \cdot \hat{\mathbf{e}}_k$ that we sought.

Observe the following properties that follow from it:

$$\left.\begin{aligned}
&\frac{\partial \hat{\mathbf{e}}_i}{\partial \hat{x}_j} \cdot \hat{\mathbf{e}}_k = 0 \quad \text{if} \quad (i, j, k) \text{ distinct,} \\
&\frac{\partial \hat{\mathbf{e}}_i}{\partial \hat{x}_j} \cdot \hat{\mathbf{e}}_k = 0 \quad \text{if } k = i, \\
&\frac{\partial \hat{\mathbf{e}}_i}{\partial \hat{x}_{\underline{i}}} \cdot \hat{\mathbf{e}}_k = -\frac{1}{h_{\underline{k}}} \frac{\partial h_i}{\partial \hat{x}_k}, \quad \text{if} \quad i \neq k \\
&\frac{\partial \hat{\mathbf{e}}_i}{\partial \hat{x}_{\underline{k}}} \cdot \hat{\mathbf{e}}_k = -\frac{1}{h_{\underline{i}}} \frac{\partial h_k}{\partial \hat{x}_i} \quad \text{if} \quad i \neq k.
\end{aligned}\right\} \tag{6.30}$$

6.3 Transformation of Basic Tensor Relations

Let $\mathbb{T}$ be a cartesian tensor field of order $N \geq 1$, defined on a region $R \subset \mathbb{E}_3$ and suppose that the points of R are regular points of $\mathbb{E}_3$ with respect to a given orthogonal curvilinear coordinate system.

The *curvilinear components* $\hat{T}_{ijk\ldots n}$ of $\mathbb{T}$ are the components of $\mathbb{T}$ in the local basis $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$. Thus,

$$\hat{T}_{ij\ldots n} = Q_{ip} Q_{jq} \ldots Q_{nr} T_{pq\ldots r}, \qquad \text{where} \qquad Q_{ip} = \hat{\mathbf{e}}_i \cdot \mathbf{e}_p. \tag{6.31}$$

6.3.1 Gradient of a scalar field

Let $\phi(\mathbf{x})$ be a scalar-valued function and let $\mathbf{v}(\mathbf{x})$ denote its gradient:

$$\mathbf{v} = \boldsymbol{\nabla}\phi \quad \text{or equivalently} \qquad v_i = \phi_{,i}.$$

The components of $\mathbf{v}$ in the two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ are related in the usual way by

$$\hat{v}_k = Q_{ki} v_i$$

and so

$$\hat{v}_k = Q_{ki} v_i = Q_{ki} \frac{\partial \phi}{\partial x_i}.$$

On using (6.22) this leads to

$$\hat{v}_k = \left(\frac{1}{h_{\underline{k}}} \frac{\partial x_i}{\partial \hat{x}_k} \right) \frac{\partial \phi}{\partial x_i} = \frac{1}{h_{\underline{k}}} \, \frac{\partial \phi}{\partial x_i} \frac{\partial x_i}{\partial \hat{x}_k},$$

so that by the chain rule

$$\hat{v}_k = \frac{1}{h_{\underline{k}}} \frac{\partial \hat{\phi}}{\partial \hat{x}_k}, \tag{6.32}$$

where we have set

$$\hat{\phi}(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \phi(x_1(\hat{x}_1, \hat{x}_2, \hat{x}_3), x_2(\hat{x}_1, \hat{x}_2, \hat{x}_3), x_3(\hat{x}_1, \hat{x}_2, \hat{x}_3)).$$

6.3.2 Gradient of a vector field

Let $\mathbf{v}(\mathbf{x})$ be a vector-valued function and let $\mathbf{W}(\mathbf{x})$ denote its gradient:

$$\mathbf{W} = \boldsymbol{\nabla}\mathbf{v} \quad \text{or equivalently} \quad W_{ij} = v_{i,j}.$$

The components of $\mathbf{W}$ and $\mathbf{v}$ in the two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ are related in the usual way by

$$\widehat{W}_{ij} = Q_{ip}Q_{jq}W_{pq}, \qquad v_p = Q_{np}\hat{v}_n,$$

and therefore

$$\widehat{W}_{ij} = Q_{ip}Q_{jq}\frac{\partial v_p}{\partial x_q} = Q_{ip}Q_{jq}\frac{\partial}{\partial x_q}(Q_{np}\hat{v}_n) = Q_{ip}Q_{jq}\frac{\partial}{\partial \hat{x}_m}(Q_{np}\hat{v}_n)\frac{\partial \hat{x}_m}{\partial x_q}.$$

Thus by (6.24)[4]

$$\widehat{W}_{ij} = Q_{ip}Q_{jq}\sum_{m=1}^{3}\frac{1}{h_m}Q_{mq}\frac{\partial}{\partial \hat{x}_m}(Q_{np}\hat{v}_n) \ .$$

Since $Q_{jq}Q_{mq} = \delta_{mj}$, this simplifies to

$$\widehat{W}_{ij} = Q_{ip}\frac{1}{h_{\underline{j}}}\frac{\partial}{\partial \hat{x}_j}(Q_{np}\hat{v}_n),$$

which, on expanding the terms in parentheses, yields

$$\widehat{W}_{ij} \ = \ \frac{1}{h_{\underline{j}}}\left\{\frac{\partial \hat{v}_i}{\partial \hat{x}_j} + Q_{ip}\frac{\partial Q_{np}}{\partial \hat{x}_j}\hat{v}_n\right\}.$$

However by (6.22)

$$Q_{ip}\frac{\partial Q_{np}}{\partial \hat{x}_j}\hat{v}_n = Q_{ip}\frac{\partial}{\partial \hat{x}_j}(\hat{\mathbf{e}}_n \cdot \mathbf{e}_p)\hat{v}_n = \hat{\mathbf{e}}_i \cdot \frac{\partial \hat{\mathbf{e}}_n}{\partial \hat{x}_j}\hat{v}_n,$$

and so

$$\widehat{W}_{ij} = \frac{1}{h_{\underline{j}}}\left\{\frac{\partial \hat{v}_i}{\partial \hat{x}_j} + \left[\hat{\mathbf{e}}_i \cdot \frac{\partial \hat{\mathbf{e}}_n}{\partial \hat{x}_j}\right]\hat{v}_n\right\}, \tag{6.33}$$

in which the coefficient in brackets is given by (6.29).

[4] We explicitly use the summation sign in this equation (and elsewhere) when an index is repeated 3 or more times, and we wish sum over it.

6.3.3 Divergence of a vector field

Let $\mathbf{v}(\mathbf{x})$ be a vector-valued function and let $\mathbf{W}(\mathbf{x}) = \boldsymbol{\nabla}\mathbf{v}(\mathbf{x})$ denote its gradient. Then

$$\text{div } \mathbf{v} = \text{trace } \mathbf{W} = v_{i,i}.$$

Therefore from (6.33), the invariance of the trace of $\mathbf{W}$, and (6.30),

$$\text{div } \mathbf{v} = \text{trace } \mathbf{W} = \text{trace } \widehat{\mathbf{W}} = \widehat{W}_{ii} = \sum_{i=1}^{3} \frac{1}{h_i} \left\{ \frac{\partial \hat{v}_i}{\partial \hat{x}_i} + \sum_{n \neq i} \frac{1}{h_n} \frac{\partial h_i}{\partial \hat{x}_n} \hat{v}_n \right\}.$$

Collecting terms involving $\hat{v}_1$, $\hat{v}_2$, and $\hat{v}_3$ alone, one has

$$\text{div } \mathbf{v} = \frac{1}{h_1} \frac{\partial \hat{v}_1}{\partial \hat{x}_1} + \frac{\hat{v}_1}{h_2 h_1} \frac{\partial h_2}{\partial \hat{x}_1} + \frac{\hat{v}_1}{h_3 h_1} \frac{\partial h_3}{\partial \hat{x}_1} + \ldots + \ldots$$

Thus

$$\text{div } \mathbf{v} = \frac{1}{h_1 h_2 h_3} \left\{ \frac{\partial}{\partial \hat{x}_1}(h_2 h_3 \hat{v}_1) + \frac{\partial}{\partial \hat{x}_2}(h_3 h_1 \hat{v}_2) + \frac{\partial}{\partial \hat{x}_3}(h_1 h_2 \hat{v}_3) \right\}. \tag{6.34}$$

6.3.4 Laplacian of a scalar field

Let $\phi(\mathbf{x})$ be a scalar-valued function. Since

$$\nabla^2 \phi = \text{div}(\text{grad } \phi) = \phi_{,kk}$$

the results from Sub-sections (6.3.1) and (6.3.3) permit us to infer that

$$\nabla^2 \phi = \frac{1}{h_1 h_2 h_3} \left\{ \frac{\partial}{\partial \hat{x}_1} \Big(\frac{h_2 h_3}{h_1} \frac{\partial \hat{\phi}}{\partial \hat{x}_1} \Big) + \frac{\partial}{\partial \hat{x}_2} \Big(\frac{h_3 h_1}{h_2} \frac{\partial \hat{\phi}}{\partial \hat{x}_2} \Big) + \frac{\partial}{\partial \hat{x}_3} \Big(\frac{h_1 h_2}{h_3} \frac{\partial \hat{\phi}}{\partial \hat{x}_3} \Big) \right\} \tag{6.35}$$

where we have set

$$\hat{\phi}(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \phi(x_1(\hat{x}_1, \hat{x}_2, \hat{x}_3), x_2(\hat{x}_1, \hat{x}_2, \hat{x}_3), x_3(\hat{x}_1, \hat{x}_2, \hat{x}_3)).$$

6.3.5 Curl of a vector field

Let $\mathbf{v}(\mathbf{x})$ be a vector-valued field and let $\mathbf{w}(\mathbf{x})$ be its curl so that

$$\mathbf{w} = \text{curl } \mathbf{v} \quad \text{or equivalently} \quad w_i = e_{ijk} v_{k,j}.$$

Let

$$\mathbf{W} = \boldsymbol{\nabla}\mathbf{v} \quad \text{or equivalently} \quad W_{ij} = v_{i,j}\ .$$

Then as we have shown in an earlier chapter

$$w_i = e_{ijk}W_{kj}, \qquad \widehat{w}_i = e_{ijk}\widehat{W}_{kj}.$$

Consequently from Sub-section 6.3.2,

$$\widehat{w}_i = \sum_{j=1}^{3} \frac{1}{h_j} e_{ijk} \left\{ \frac{\partial \hat{v}_k}{\partial \hat{x}_j} + \left[\hat{\mathbf{e}}_k \cdot \frac{\partial \hat{\mathbf{e}}_n}{\partial \hat{x}_j} \right] \hat{v}_n \right\}.$$

By (6.30), the second term within the braces sums out to zero unless $n = j$. Thus, using the second of (6.30), one arrives at

$$\widehat{w}_i = \sum_{j,k=1}^{3} \frac{1}{h_j} e_{ijk} \left\{ \frac{\partial \hat{v}_k}{\partial \hat{x}_j} - \frac{1}{h_k}\frac{\partial h_j}{\partial \hat{x}_k} \hat{v}_j \right\}.$$

This yields

$$\begin{aligned}
\widehat{w}_1 &= \frac{1}{h_2h_3}\left\{ \frac{\partial}{\partial \hat{x}_2}(h_3\hat{v}_3) - \frac{\partial}{\partial \hat{x}_3}(h_2\hat{v}_2) \right\}, \\
\widehat{w}_2 &= \frac{1}{h_3h_1}\left\{ \frac{\partial}{\partial \hat{x}_3}(h_1\hat{v}_1) - \frac{\partial}{\partial \hat{x}_1}(h_3\hat{v}_3) \right\}, \qquad (6.36) \\
\widehat{w}_3 &= \frac{1}{h_1h_2}\left\{ \frac{\partial}{\partial \hat{x}_1}(h_2\hat{v}_2) - \frac{\partial}{\partial \hat{x}_2}(h_1\hat{v}_1) \right\}.
\end{aligned}$$

6.3.6 Divergence of a symmetric 2-tensor field

Let $\mathbf{S}(\mathbf{x})$ be a symmetric 2-tensor field and let $\mathbf{v}(x)$ denote its divergence:

$$\mathbf{v} = \text{div}\ \mathbf{S}, \quad \mathbf{S} = \mathbf{S}^T, \quad \text{or equivalently} \quad v_i = S_{ij,j}, \quad S_{ij} = S_{ji}.$$

The components of $\mathbf{v}$ and $\mathbf{S}$ in the two bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ are related in the usual way by

$$\hat{v}_i = Q_{ip}v_p, \quad S_{ij} = Q_{mi}Q_{nj}\hat{S}_{mn},$$

and consequently

$$\hat{v}_i = Q_{ip}S_{pj,j} = Q_{ip}\frac{\partial}{\partial x_j}(Q_{mp}Q_{nj}\hat{S}_{mn}) = Q_{ip}\frac{\partial}{\partial \hat{x}_k}(Q_{mp}Q_{nj}\hat{S}_{mn})\frac{\partial \hat{x}_k}{\partial x_j}.$$

By using (6.24), the orthogonality of the matrix $[Q]$, (6.30) and $\hat{S}_{ij} = \hat{S}_{ji}$ we obtain

$$\begin{aligned}\hat{v}_1 &= \frac{1}{h_1h_2h_3}\left\{\frac{\partial}{\partial\hat{x}_1}(h_2h_3\hat{S}_{11}) + \frac{\partial}{\partial\hat{x}_2}(h_3h_1\hat{S}_{12}) + \frac{\partial}{\partial\hat{x}_3}(h_1h_2\hat{S}_{13})\right\}\\ &+ \frac{1}{h_1h_2}\frac{\partial h_1}{\partial\hat{x}_2}\hat{S}_{12} + \frac{1}{h_1h_3}\frac{\partial h_1}{\partial\hat{x}_3}\hat{S}_{13} - \frac{1}{h_1h_2}\frac{\partial h_2}{\partial\hat{x}_1}\hat{S}_{22} - \frac{1}{h_1h_3}\frac{\partial h_3}{\partial\hat{x}_1}\hat{S}_{33},\end{aligned} \tag{6.37}$$

with analogous expressions for $\hat{v}_2$ and $\hat{v}_3$.

Equations (6.32) - (6.37) provide the fundamental expressions for the basic tensor-analytic quantities that we will need. Observe that they reduce to their classical rectangular cartesian forms in the special case $x_i = \hat{x}_i$ (in which case $h_1 = h_2 = h_3 = 1$).

6.3.7 Differential elements of volume

When evaluating a volume integral over a region $\mathcal{D}$, we sometimes find it convenient to transform it from the form

$$\int_{\mathcal{D}} \Big(\ldots\Big) dx_1dx_2dx_3 \quad \text{into an equivalent expression of the form} \quad \int_{\mathcal{D}'} \Big(\ldots\Big) d\hat{x}_1d\hat{x}_2d\hat{x}_3.$$

In order to do this we must relate $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$ to $\mathrm{d}\hat{x}_1\mathrm{d}\hat{x}_2\mathrm{d}\hat{x}_3$. By (6.22),

$$\det[Q] = \frac{1}{h_1h_2h_3}\det[J].$$

However since $[Q]$ is a proper orthogonal matrix its determinant takes the value $+1$. Therefore

$$\det[J] = h_1h_2h_3$$

and so the basic relation $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = \det[J]\ \mathrm{d}\hat{x}_1\mathrm{d}\hat{x}_2\mathrm{d}\hat{x}_3$ leads to

$$\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = h_1h_2h_3\ \mathrm{d}\hat{x}_1\mathrm{d}\hat{x}_2\mathrm{d}\hat{x}_3. \tag{6.38}$$

6.3.8 Differential elements of area

Let $\mathbf{d\hat{A}}_1$ denote a differential element of (vector) area on a $\hat{x}_1$-coordinate surface so that $\mathbf{d\hat{A}}_1 = (d\hat{x}_2\,\partial\mathbf{x}/\partial\hat{x}_2) \times (\mathrm{d}\hat{x}_3\,\partial\mathbf{x}/\partial\hat{x}_3)$. In view of (6.21) this leads to $\mathbf{d\hat{A}_1} = (d\hat{x}_2\,h_2\,\hat{\mathbf{e}}_2) \times (\mathrm{d}\hat{x}_3\,h_3\,\hat{\mathbf{e}}_3) = h_2h_3\mathrm{d}\hat{x}_2\mathrm{d}\hat{x}_3\ \hat{\mathbf{e}}_1$. Thus the differential elements of (scalar) area on the $\hat{x}_1$-, $\hat{x}_2$- and $\hat{x}_3$-coordinate surfaces are given by

$$\mathrm{d}\hat{A}_1 = h_2h_3\mathrm{d}\hat{x}_2\mathrm{d}\hat{x}_3, \qquad \mathrm{d}\hat{A}_2 = h_3h_1\mathrm{d}\hat{x}_3\mathrm{d}\hat{x}_1, \qquad \mathrm{d}\hat{A}_3 = h_1h_2\mathrm{d}\hat{x}_1\mathrm{d}\hat{x}_2, \tag{6.39}$$

respectively.

6.4 Some Examples of Orthogonal Curvilinear Coordinate Systems

Circular Cylindrical Coordinates (r, θ, z):

$$\left.\begin{array}{c} x_1 = r\cos\theta, \qquad x_2 = r\sin\theta, \qquad x_3 = z; \\ \text{for all } (r,\theta,z) \in [0,\infty)\times[0,2\pi)\times(-\infty,\infty); \\ h_r = 1, \quad h_\theta = r, \quad h_z = 1. \end{array}\right\} \tag{6.40}$$

Spherical Coordinates (r, θ, ϕ):

$$\left.\begin{array}{c} x_1 = r\sin\theta\cos\phi, \qquad x_2 = r\sin\theta\sin\phi, \qquad x_3 = r\cos\theta; \\ \text{for all } (r,\theta,\phi) \in [0,\infty)\times[0,2\pi)\times(-\pi,\pi]; \\ h_r = 1, \quad h_\theta = r, \quad h_\phi = r\sin\theta\ . \end{array}\right\} \tag{6.41}$$

Elliptical Cylindrical Coordinates (ξ, η, z):

$$\left.\begin{array}{c} x_1 = a\ \cosh\xi\cos\eta, \qquad x_2 = a\ \sinh\xi\sin\eta, \qquad x_3 = z; \\ \text{for all } (\xi,\eta,z) \in [0,\infty)\times(-\pi,\pi]\times(-\infty,\infty); \\ h_\xi = h_\eta = a\sqrt{\sinh^2\xi + \sin^2\eta}, \quad h_z = 1\ . \end{array}\right\} \tag{6.42}$$

Parabolic Cylindrical Coordinates (u, v, w):

$$\left.\begin{array}{c} x_1 = \tfrac{1}{2}(u^2 - v^2), \qquad x_2 = uv, \qquad x_3 = w; \\ \text{for all } (u,v,w) \in (-\infty,\infty)\times[0,\infty)\times(-\infty,\infty); \\ h_u = h_v = \sqrt{u^2+v^2}, \quad h_z = 1\ . \end{array}\right\} \tag{6.43}$$

6.5 Worked Examples.

Example 6.1: Let $\mathbf{E}(\mathbf{x})$ be a symmetric 2-tensor field that is related to a vector field $\mathbf{u}(\mathbf{x})$ through

$$\mathbf{E} = \frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\right).$$

In a cartesian coordinate system this can be written equivalently as

$$E_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right).$$

Establish the analogous formulas in a general orthogonal curvilinear coordinate system.

<u>*Solution:*</u> Using the result from Sub-section 6.3.2 and the formulas for $\hat{\mathbf{e}}_k \cdot (\partial \hat{\mathbf{e}}_i / \partial \hat{x}_j$ one finds after elementary simplification that

$$\left.\begin{array}{lcl} \widehat{E}_{11} & = & \dfrac{1}{h_1}\dfrac{\partial \hat{u}_1}{\partial \hat{x}_1} + \dfrac{1}{h_1 h_2}\dfrac{\partial h_1}{\partial \hat{x}_2}\hat{u}_2 + \dfrac{1}{h_1 h_3}\dfrac{\partial h_1}{\partial \hat{x}_3}\hat{u}_3, \qquad E_{22} = \ldots, \quad E_{33} = \ldots, \\ \widehat{E}_{12} & = & \widehat{E}_{21} = \dfrac{1}{2}\left\{\dfrac{h_1}{h_2}\dfrac{\partial}{\partial \hat{x}_2}\left(\dfrac{\hat{u}_1}{h_1}\right) + \dfrac{h_2}{h_1}\dfrac{\partial}{\partial \hat{x}_1}\left(\dfrac{\hat{u}_2}{h_2}\right)\right\}, \quad E_{23} = \ldots, \quad E_{31} = \ldots, \end{array}\right\} \qquad (i)$$

<u>*Example 6.2:*</u> Consider a symmetric 2-tensor field $\mathbf{S}(\mathbf{x})$ and a vector field $\mathbf{b}(\mathbf{x})$ that satisfy the equation

$$\text{div}\ \mathbf{S} + \mathbf{b} = \mathbf{o}.$$

In a cartesian coordinate system this can be written equivalently as

$$\frac{\partial S_{ij}}{\partial x_j} + b_i = 0.$$

Establish the analogous formulas in a general orthogonal curvilinear coordinate system.

<u>*Solution:*</u> From the results in Sub-section 6.3.6 we have

$$\begin{array}{l} \dfrac{1}{h_1 h_2 h_3}\left\{\dfrac{\partial}{\partial \hat{x}_1}(h_2 h_3 \widehat{S}_{11}) + \dfrac{\partial}{\partial \hat{x}_2}(h_3 h_1 \widehat{S}_{12}) + \dfrac{\partial}{\partial \hat{x}_3}(h_1 h_2 \widehat{S}_{13})\right\} \\ \qquad + \quad \dfrac{1}{h_1 h_2}\dfrac{\partial h_1}{\partial \hat{x}_2}\widehat{S}_{12} + \dfrac{1}{h_1 h_3}\dfrac{\partial h_1}{\partial \hat{x}_3}\widehat{S}_{13} - \dfrac{1}{h_1 h_2}\dfrac{\partial h_2}{\partial \hat{x}_1}\widehat{S}_{22} - \dfrac{1}{h_1 h_3}\dfrac{\partial h_3}{\partial \hat{x}_1}\widehat{S}_{33} + \hat{b}_1 = 0, \end{array} \qquad (i)$$

.........etc.

where

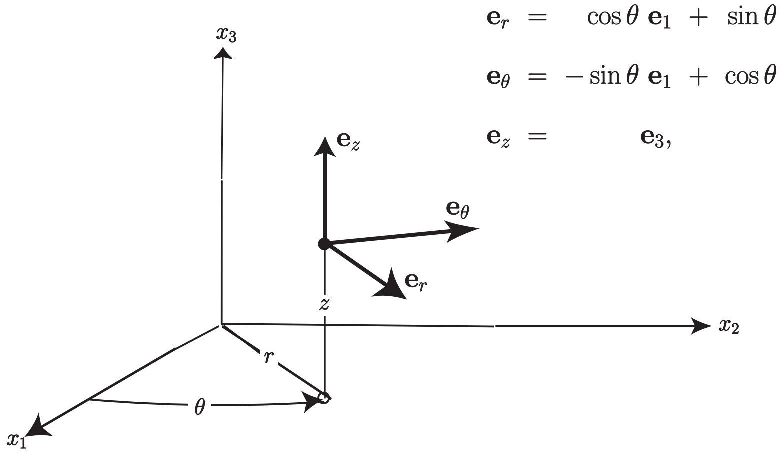
$$\hat{b}_i = Q_{ip} b_p$$

<u>*Example 6.3:*</u> Consider *circular cylindrical coordinates* $(\hat{x}_1, \hat{x}_2, \hat{x}_3) = (r, \theta, z)$ which are related to (x_1, x_2, x_3) through

$$\left.\begin{array}{c} x_1 = r\cos\theta, \quad x_2 = r\sin\theta, \quad x_3 = z, \\ \\ 0 \le r < \infty, \quad 0 \le \theta < 2\pi, \quad -\infty < z < \infty. \end{array}\right\}$$

Let $f(\mathbf{x})$ be a scalar-valued field, $\mathbf{u}(\mathbf{x})$ a vector-valued field, and $\mathbf{S}(\mathbf{x})$ a symmetric 2-tensor field. Express the following quanties,

(a) grad f



$$\left.\begin{aligned} \mathbf{e}_r &= \cos\theta\,\mathbf{e}_1 + \sin\theta\,\mathbf{e}_2, \\ \mathbf{e}_\theta &= -\sin\theta\,\mathbf{e}_1 + \cos\theta\,\mathbf{e}_2, \\ \mathbf{e}_z &= \mathbf{e}_3, \end{aligned}\right\}$$

Figure 6.3: Cylindrical coordinates (r, θ, z) and the associated local curvilinear orthonormal basis $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z\}$.

(b) $\nabla^2 f$

(c) div $\mathbf{u}$

(d) curl $\mathbf{u}$

(e) $\frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\right)$ and

(f) div $\mathbf{S}$

in this coordinate system.

Solution: We simply need to specialize the basic results established in Section 6.3.

In the present case we have

$$(\hat{x}_1, \hat{x}_2, \hat{x}_3) = (r, \theta, z) \tag{i}$$

and the coordinate mapping (6.5) takes the particular form

$$x_1 = r\cos\theta, \qquad x_2 = r\sin\theta, \qquad x_3 = z. \tag{ii}$$

The matrix $[\partial x_i/\partial \hat{x}_j]$ therefore specializes to

$$\begin{pmatrix} \partial x_1/\partial r & \partial x_1/\partial\theta & \partial x_1/\partial z \\ \partial x_2/\partial r & \partial x_2/\partial\theta & \partial x_2/\partial z \\ \partial x_3/\partial r & \partial x_3/\partial\theta & \partial x_3/\partial z \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and the scale moduli are

$$\begin{aligned} h_r &= \sqrt{\left(\frac{\partial x_1}{\partial r}\right)^2 + \left(\frac{\partial x_2}{\partial r}\right)^2 + \left(\frac{\partial x_3}{\partial r}\right)^2} &= 1,\\ h_\theta &= \sqrt{\left(\frac{\partial x_1}{\partial \theta}\right)^2 + \left(\frac{\partial x_2}{\partial \theta}\right)^2 + \left(\frac{\partial x_3}{\partial \theta}\right)^2} &= r,\\ h_z &= \sqrt{\left(\frac{\partial x_1}{\partial z}\right)^2 + \left(\frac{\partial x_2}{\partial z}\right)^2 + \left(\frac{\partial x_3}{\partial z}\right)^2} &= 1. \end{aligned} \tag{iii}$$

We use the natural notation

$$(u_r, u_\theta, u_z) = (\hat{u}_1, \hat{u}_2, \hat{u}_3) \tag{iv}$$

for the components of a vector field, and

$$(S_{rr}, S_{r\theta}, \ldots) = (\hat{S}_{11}, \hat{S}_{12}, \ldots) \tag{v}$$

for the components of a 2-tensor field, and

$$(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z) = (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3) \tag{vi}$$

for the unit vectors associated with the local cylindrical coordinate system.

From (ii),

$$\mathbf{x} = (r\cos\theta)\mathbf{e}_1 + (r\sin\theta)\mathbf{e}_2 + (z)\mathbf{e}_3,$$

and therefore on using (6.21) and (iii) we obtain the following expressions for the unit vectors associated with the local cylindrical coordinate system:

$$\left.\begin{aligned} \mathbf{e}_r &= \cos\theta\ \mathbf{e}_1 + \sin\theta\ \mathbf{e}_2,\\ \mathbf{e}_\theta &= -\sin\theta\ \mathbf{e}_1 + \cos\theta\ \mathbf{e}_2,\\ \mathbf{e}_z &= \mathbf{e}_3, \end{aligned}\right\}$$

which, in this case, could have been obtained geometrically from Figure 6.3.

(*a*) Substituting (i) and (iii) into (6.32) gives

$$\boldsymbol{\nabla} f = \left(\frac{\partial \widehat{f}}{\partial r}\right)\mathbf{e}_r + \left(\frac{1}{r}\frac{\partial \widehat{f}}{\partial \theta}\right)\mathbf{e}_\theta + \left(\frac{\partial \widehat{f}}{\partial z}\right)\mathbf{e}_z$$

where we have set $\widehat{f}(r,\theta,z) = f(x_1,x_2,x_3)$.

(*b*) Substituting (i) and (iii) into (6.35) gives

$$\nabla^2 \widehat{f} = \frac{\partial^2 \widehat{f}}{\partial r^2} + \frac{1}{r}\frac{\partial \widehat{f}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \widehat{f}}{\partial \theta^2} + \frac{\partial^2 \widehat{f}}{\partial z^2}$$

where we have set $\widehat{f}(r,\theta,z) = f(x_1,x_2,x_3)$.

(*c*) Substituting (i) and (iii) into (6.34) gives

$$\text{div } \mathbf{u} = \frac{\partial u_r}{\partial r} + \frac{1}{r}u_r + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z}$$

(*d*) Substituting (i) and (iii) into (6.36) gives

$$\text{curl } \mathbf{u} = \left(\frac{1}{r}\frac{\partial u_z}{\partial \theta} - \frac{\partial u_\theta}{\partial z}\right)\mathbf{e}_r + \left(\frac{\partial u_r}{\partial z} - \frac{\partial u_z}{\partial r}\right)\mathbf{e}_\theta + \left(\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{r} - \frac{1}{r}\frac{\partial u_r}{\partial \theta}\right)\mathbf{e}_z$$

(*e*) Set $\mathbf{E}$=$(1/2)(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T)$. Substituting (i) and (iii) into (6.33) enables us to calculate $\boldsymbol{\nabla}\mathbf{u}$ whence we can calculate $\mathbf{E}$. Writing the cylindrical components $\hat{E}_{ij}$ of $\mathbf{E}$ as

$$(E_{rr}, E_{r\theta}, E_{rz}, \ldots) = (\hat{E}_{11}, \hat{E}_{12}, \hat{E}_{13} \ldots),$$

one finds

$$\left.\begin{aligned}
E_{rr} &= \frac{\partial u_r}{\partial r}, \\
E_{\theta\theta} &= \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{u_r}{r}, \\
E_{zz} &= \frac{\partial u_z}{\partial z}, \\
E_{r\theta} &= \frac{1}{2}\left(\frac{1}{r}\frac{\partial u_r}{\partial \theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right), \\
E_{\theta z} &= \frac{1}{2}\left(\frac{\partial u_\theta}{\partial z} + \frac{1}{r}\frac{\partial u_z}{\partial \theta}\right), \\
E_{zr} &= \frac{1}{2}\left(\frac{\partial u_z}{\partial r} + \frac{\partial u_r}{\partial z}\right).
\end{aligned}\right\}$$

Alternatively these could have been obtained from the results of Example 6.1.

(*f*) Finally, substituting (i) and (iii) into (6.37) gives

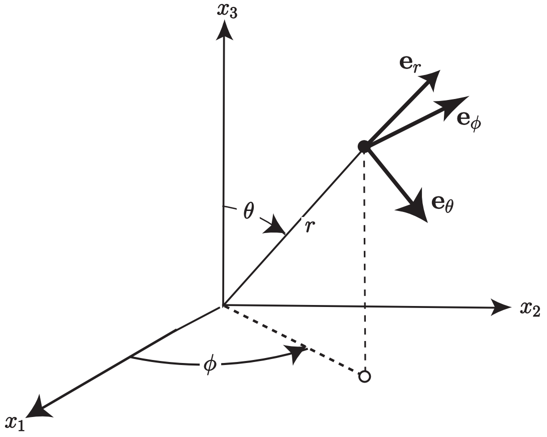
$$\begin{aligned}
\text{div}\,\mathbf{S} \;=\; & \left(\frac{\partial S_{rr}}{\partial r} + \frac{1}{r}\frac{\partial S_{r\theta}}{\partial \theta} + \frac{\partial S_{rz}}{\partial z} + \frac{S_{rr} - S_{\theta\theta}}{r}\right)\mathbf{e}_r \\
+ \; & \left(\frac{\partial S_{r\theta}}{\partial r} + \frac{1}{r}\frac{\partial S_{\theta\theta}}{\partial \theta} + \frac{\partial S_{\theta z}}{\partial z} + \frac{2S_{r\theta}}{r}\right)\mathbf{e}_\theta \\
+ \; & \left(\frac{\partial S_{zr}}{\partial r} + \frac{1}{r}\frac{\partial S_{z\theta}}{\partial \theta} + \frac{\partial S_{zz}}{\partial z} + \frac{S_{zr}}{r}\right)\mathbf{e}_z
\end{aligned}$$

Alternatively these could have been obtained from the results of Example 6.2.

Example 6.4: Consider *spherical coordinates* $(\hat{x}_1, \hat{x}_2, \hat{x}_3) = (r, \theta, \phi)$ which are related to (x_1, x_2, x_3) through

$$\left.\begin{array}{c}
x_1 = r\sin\theta\cos\phi, \quad x_2 = r\sin\theta\sin\phi, \quad x_3 = r\cos\theta, \\[1ex]
0 \le r < \infty, \quad 0 \le \theta \le \pi, \quad 0 \le \phi < 2\pi.
\end{array}\right\}$$

Let $f(\mathbf{x})$ be a scalar-valued field, $\mathbf{u}(\mathbf{x})$ a vector-valued field, and $\mathbf{S}(\mathbf{x})$ a symmetric 2-tensor field. Express the following quanties,



$$\left.\begin{aligned}
\mathbf{e}_r &= (\sin\theta\cos\phi)\ \mathbf{e}_1 + (\sin\theta\sin\phi)\ \mathbf{e}_2 + \cos\theta\ \mathbf{e}_3,\\
\mathbf{e}_\theta &= (\cos\theta\cos\phi)\ \mathbf{e}_1 + (\cos\theta\sin\phi)\ \mathbf{e}_2 - \sin\theta\ \mathbf{e}_3,\\
\mathbf{e}_\phi &= -\sin\phi\ \mathbf{e}_1 + \cos\phi\ \mathbf{e}_2,
\end{aligned}\right\}$$

Figure 6.4: Spherical coordinates (r, θ, ϕ) and the associated local curvilinear orthonormal basis $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi\}$.

(a) grad f

(b) div $\mathbf{u}$

(c) $\nabla^2 f$

(d) curl $\mathbf{u}$

(e) $\frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\right)$ and

(f) div $\mathbf{S}$

in this coordinate system.

Solution: We simply need to specialize the basic results established in Section 6.3.

In the present case we have

$$(\hat{x}_1, \hat{x}_2, \hat{x}_3) = (r, \theta, \phi), \tag{i}$$

and the coordinate mapping (6.5) takes the particular form

$$x_1 = r\sin\theta\cos\phi, \qquad x_2 = r\sin\theta\sin\phi, \qquad x_3 = r\cos\theta. \tag{ii}$$

The matrix $[\partial x_i/\partial\hat{x}_j]$ therefore specializes to

$$\begin{pmatrix} \partial x_1/\partial r & \partial x_1/\partial\theta & \partial x_1/\partial\phi \\ \partial x_2/\partial r & \partial x_2/\partial\theta & \partial x_2/\partial\phi \\ \partial x_3/\partial r & \partial x_3/\partial\theta & \partial x_3/\partial\phi \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}$$

and the scale moduli are

$$\begin{aligned} h_r &= \sqrt{\left(\frac{\partial x_1}{\partial r}\right)^2 + \left(\frac{\partial x_2}{\partial r}\right)^2 + \left(\frac{\partial x_3}{\partial r}\right)^2} &= 1, \\ h_\theta &= \sqrt{\left(\frac{\partial x_1}{\partial \theta}\right)^2 + \left(\frac{\partial x_2}{\partial \theta}\right)^2 + \left(\frac{\partial x_3}{\partial \theta}\right)^2} &= r, \\ h_\phi &= \sqrt{\left(\frac{\partial x_1}{\partial \phi}\right)^2 + \left(\frac{\partial x_2}{\partial \phi}\right)^2 + \left(\frac{\partial x_3}{\partial \phi}\right)^2} &= r\sin\theta. \end{aligned} \tag{iii}$$

We use the natural notation

$$(u_r, u_\theta, u_\phi) = (\hat{u}_1, \hat{u}_2, \hat{u}_3) \tag{iv}$$

for the components of a vector field,

$$(S_{rr}, S_{r\theta}, S_{r\phi} \ldots) = (\hat{S}_{11}, \hat{S}_{12}, \hat{S}_{13} \ldots) \tag{v}$$

for the components of a 2-tensor field, and

$$(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi) = (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3) \tag{vi}$$

for the unit vectors associated with the local spherical coordinate system.

From (ii),

$$\mathbf{x} = (r\sin\theta\cos\phi)\mathbf{e}_1 + (r\sin\theta\sin\phi)\mathbf{e}_2 + (r\cos\theta)\mathbf{e}_3,$$

and therefore on using (6.21) and (iii) we obtain the following expressions for the unit vectors associated with the local spherical coordinate system:

$$\left.\begin{aligned} \mathbf{e}_r &= (\sin\theta\cos\phi)\ \mathbf{e}_1 + (\sin\theta\sin\phi)\ \mathbf{e}_2 + \cos\theta\ \mathbf{e}_3, \\ \mathbf{e}_\theta &= (\cos\theta\cos\phi)\ \mathbf{e}_1 + (\cos\theta\sin\phi)\ \mathbf{e}_2 - \sin\theta\ \mathbf{e}_3, \\ \mathbf{e}_\phi &= -\sin\phi\ \mathbf{e}_1 + \cos\phi\ \mathbf{e}_2, \end{aligned}\right\}$$

which, in this case, could have been obtained geometrically from Figure 6.4.

(a) Substituting (i) and (iii) into (6.32) gives

$$\boldsymbol{\nabla} f = \left(\frac{\partial \widehat{f}}{\partial r}\right)\mathbf{e}_r + \left(\frac{1}{r}\frac{\partial \widehat{f}}{\partial \theta}\right)\mathbf{e}_\theta + \left(\frac{1}{r\sin\theta}\frac{\partial \widehat{f}}{\partial \phi}\right)\mathbf{e}_\phi.$$

where we have set $\widehat{f}(r,\theta,\phi) = f(x_1, x_2, x_3)$.

(b) Substituting (i) and (iii) into (6.35) gives

$$\boldsymbol{\nabla}^2 f = \frac{\partial^2 \widehat{f}}{\partial r^2} + \frac{2}{r}\frac{\partial \widehat{f}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \widehat{f}}{\partial \theta^2} + \frac{1}{r^2}\cot\theta\frac{\partial \widehat{f}}{\partial \theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 \widehat{f}}{\partial \phi^2}$$

where we have set $\widehat{f}(r,\theta,\phi) = f(x_1, x_2, x_3)$.

(c) Substituting (i) and (iii) into (6.34) gives

$$\text{div } \mathbf{u} = \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}(r^2\sin\theta\ u_r) + \frac{\partial}{\partial\theta}(r\sin\theta u_\theta) + \frac{\partial}{\partial\phi}(ru_\phi)\right] .$$

(d) Substituting (i) and (iii) into (6.36) gives

$$\begin{aligned}\text{curl } \mathbf{u} &= \left(\frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial\theta}(r\sin\theta v_\phi) - \frac{\partial}{\partial\phi}(rv_\theta)\right]\right)\mathbf{e}_r + \left(\frac{1}{r\sin\theta}\left[\frac{\partial v_r}{\partial\phi} - \frac{\partial}{\partial r}(r\sin\theta v_\phi)\right]\right)\mathbf{e}_\theta \\ &+ \left(\frac{1}{r}\left[\frac{\partial}{\partial r}(rv_\theta) - \frac{\partial v_r}{\partial\theta}\right]\right)\mathbf{e}_\phi \ .\end{aligned}$$

(e) Set $\mathbf{E}$=(1/2)($\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T$). We substitute (i) and (iii) into (6.33) to calculate $\boldsymbol{\nabla}\mathbf{u}$ from which one can calculate $\mathbf{E}$. Writing the spherical components $\widehat{E}_{ij}$ of $\mathbf{E}$ as

$$(E_{rr}, E_{r\theta}, E_{r\phi}, \ldots) = (\widehat{E}_{11}, \widehat{E}_{12}, \widehat{E}_{13}\ldots),$$

one finds

$$\left.\begin{aligned}E_{rr} &= \frac{\partial u_r}{\partial r}, \\ E_{\theta\theta} &= \frac{1}{r}\frac{\partial u_\theta}{\partial\theta} + \frac{u_r}{r}, \\ E_{\phi\phi} &= \frac{1}{r\sin\theta}\frac{\partial u_\phi}{\partial\phi} + \frac{u_r}{r} + \frac{\cot\theta}{r}u_\theta, \\ E_{r\theta} &= \frac{1}{2}\left(\frac{1}{r}\frac{\partial u_r}{\partial\theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right), \\ E_{\theta\phi} &= \frac{1}{2}\left(\frac{1}{r\sin\theta}\frac{\partial u_\theta}{\partial\phi} + \frac{1}{r}\frac{\partial u_\phi}{\partial\theta} - \frac{\cot\theta}{r}u_\phi\right), \\ E_{\phi r} &= \frac{1}{2}\left(\frac{1}{r\sin\theta}\frac{\partial u_r}{\partial\phi} + \frac{\partial u_\phi}{\partial r} - \frac{u_\phi}{r}\right),\end{aligned}\right\}$$

Alternatively these could have been obtained from the results of Example 6.1.

(f) Finally substituting (i) and (iii) into (6.37) gives

$$\begin{aligned}\text{div } \mathbf{S} &= \left(\frac{\partial S_{rr}}{\partial r} + \frac{1}{r}\frac{\partial S_{r\theta}}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial S_{r\phi}}{\partial\phi} + \frac{1}{r}[2S_{rr} - S_{\theta\theta} - S_{\phi\phi} + \cot\theta S_{r\theta}]\right)\mathbf{e}_r \\ &+ \left(\frac{\partial S_{r\theta}}{\partial r} + \frac{1}{r}\frac{\partial S_{\theta\theta}}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial S_{\theta\phi}}{\partial\phi} + \frac{1}{r}[3S_{r\theta} + \cot\theta(S_{\theta\theta} - S_{\phi\phi})]\right)\mathbf{e}_\theta \\ &+ \left(\frac{\partial S_{r\phi}}{\partial r} + \frac{1}{r}\frac{\partial S_{\theta\phi}}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial S_{\phi\phi}}{\partial\phi} + \frac{1}{r}[3S_{r\phi} + 2\cot\theta S_{\theta\phi}]\right)\mathbf{e}_\phi\end{aligned}$$

Alternatively these could have been obtained from the results of Example 6.2.

Example 6.5: Show that the matrix $[Q]$ defined by (6.22) is a proper orthogonal matrix.

Proof: From (6.22),

$$Q_{ij} = \frac{1}{h_{\underline{i}}}\ \frac{\partial x_j}{\partial\hat{x}_i},$$

and therefore

$$Q_{ik}Q_{jk} = \frac{1}{h_{\underline{i}}h_{\underline{j}}}\,\frac{\partial x_k}{\partial \hat{x}_i}\,\frac{\partial x_k}{\partial \hat{x}_j} = \frac{1}{h_{\underline{i}}h_{\underline{j}}} g_{ij} = \delta_{ij},$$

where in the penultimate step we have used (6.14) and in the ultimate step we have used (6.16). Thus $[Q]$ is an orthogonal matrix. Next, from (6.22) and (6.7),

$$Q_{ij} = \frac{1}{h_{\underline{i}}}\,\frac{\partial x_j}{\partial \hat{x}_i} \;=\; \frac{1}{h_{\underline{i}}}\,J_{ji}$$

where $J_{ij} = \partial x_i/\partial \hat{x}_j$ are the elements of the Jacobian matrix. Thus

$$\det[Q] = \frac{1}{h_1h_2h_3}\det[J] > 0$$

where the inequality is a consequence of the inequalities in (6.8) and (6.17). Hence $[Q]$ is proper orthogonal.

References

1. H. Reismann and P.S. Pawlik, *Elasticity: Theory and Applications*, Wiley, 1980.
2. L.A. Segel, *Mathematics Applied to Continuum Mechanics*, Dover, New York, 1987.
3. E. Sternberg, *(Unpublished) Lecture Notes for AM 135: Elasticity*, California Institute of Technology, Pasadena, California, 1976.

Chapter 7

Calculus of Variations

7.1 Introduction.

Numerous problems in physics can be formulated as mathematical problems in optimization. For example in optics, Fermat's principle states that the path taken by a ray of light in propagating from one point to another is the path that minimizes the travel time. Most equilibrium theories of mechanics involve finding a configuration of a system that minimizes its energy. For example a heavy cable that hangs under gravity between two fixed pegs adopts the shape that, from among all possible shapes, minimizes the gravitational potential energy of the system. Or, if we subject a straight beam to a compressive load, its deformed configuration is the shape which minimizes the total energy of the system. Depending on the load, the energy minimizing configuration may be straight or bent (buckled). If we dip a (non-planar) wire loop into soapy water, the soap film that forms across the loop is the one that minimizes the surface energy (which under most circumstances equals minimizing the surface area of the soap film). Another common problem occurs in geodesics where, given some surface and two points on it, we want to find the path of shortest distance joining those two points which lies entirely on the given surface.

In each of these problem we have a scalar-valued quantity F such as energy or time that depends on a function ϕ such as the shape or path, and we want to find the function ϕ that minimizes the quantity F of interest. Note that the scalar-valued function F is defined on a *set of functions*. One refers to F as a *functional* and writes $F\{\phi\}$.

As a specific example, consider the so-called Brachistochrone Problem. We are given two points $(0,0)$ and $(1,h)$ in the x,y-plane, with $h>0$, that are to be joined by a smooth wire. A bead is released from rest from the point $(0,0)$ and slides along the wire due to gravity. For what shape of wire is the time of travel from $(0,0)$ to $(1,h)$ least?

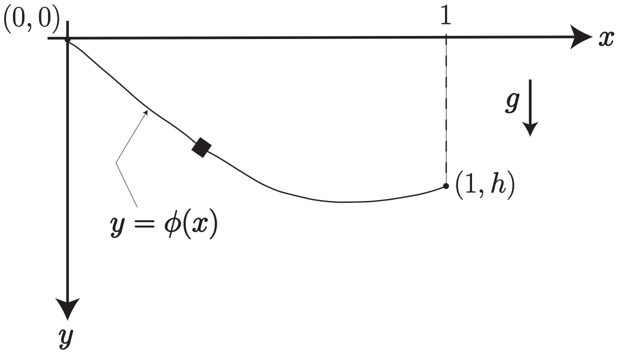


Figure 7.1: Curve joining $(0,0)$ to $(1,h)$ along which a bead slides under gravity.

In order to formulate this problem, let $y=\phi(x), 0\leq x\leq 1$, describe a generic curve joining $(0,0)$ to $(1,h)$. Let $s(t)$ denote the distance traveled by the bead along the wire at time t so that $v(t)=ds/dt$ is its corresponding speed. The travel time of the bead is

$$T=\int\frac{ds}{v}$$

where the integral is taken along the entire path. In the question posed to us, we are to find the curve, i.e. the function $\phi(x)$, which makes T a minimum. Since we are to minimize T by varying ϕ, it is natural to first rewrite the formula for T in a form that explicitly displays its dependency on ϕ.

Note first, that by elementary calculus, the arc length ds is related to dx by

$$ds=\sqrt{dx^2+dy^2}=\sqrt{1+(dy/dx)^2}\,dx=\sqrt{1+(\phi')^2}\,dx$$

and so we can write

$$T=\int_0^1\frac{\sqrt{1+(\phi')^2}}{v}\,dx.$$

Next, we wish to express the speed v in terms of ϕ. If $(x(t),y(t))$ denote the coordinates of the bead at time t, the conservation of energy tells us that the sum of the potential and kinetic energies does not vary with time:

$$-mg\phi(x(t))+\frac{1}{2}mv^2(t)=0,$$

where the right hand side is the total energy at the initial instant. Solving this for v gives

$$v = \sqrt{2g\phi}.$$

Finally, substituting this back into the formula for the travel time gives

$$T\{\phi\} = \int_0^1 \sqrt{\frac{1+(\phi')^2}{2g\phi}}\, dx. \tag{7.1}$$

Given a curve characterized by $y = \phi(x)$, this formula gives the corresponding travel time for the bead. Our task is to find, from among all such curves, the one that minimizes $T\{\phi\}$.

This minimization takes place over a set of functions ϕ. In order to complete the formulation of the problem, we should carefully characterize this set of "admissible functions" (or "test functions"). A generic curve is described by $y = \phi(x)$, $0 \leq x \leq 1$. Since we are only interested in curves that pass through the points $(0,0)$ and $(1,h)$ we must require that $\phi(0) = 0, \phi(1) = h$. Finally, for analytical reasons we only consider curves that are continuous and have a continous slope, i.e. ϕ and ϕ' are both continuous on $[0,1]$. Thus the set A of *admissible functions* that we wish to consider is

$$\mathsf{A} = \left\{\phi(\cdot) \,\middle|\, \phi : [0,1] \to \mathbb{R},\ \phi \in C^1[0,1],\ \phi(0) = 0,\ \phi(1) = h\right\}. \tag{7.2}$$

Our task is to minimize $T\{\phi\}$ over the set A.

Remark: Since the shortest distance between two points is given by the straight line that joins them, it is natural to wonder whether a straight line is also the curve that gives the minimum travel time. To investigate this, consider (a) a straight line, and (b) a circular arc, that joins $(0,0)$ to $(1,h)$. Use (7.1) to calculate the travel time for each of these paths and show that the straight line is not the path that gives the least travel time.

Remark: One can consider various variants of the Brachistochrone Problem. For example, the length of the curve joining the two points might be prescribed, in which case the minimization is to be carried out subject to the constraint that the length is given. Or perhaps the position of the left hand end might be prescribed as above, but the right hand end of the wire might be allowed to lie anywhere on the vertical line through $x = 1$. Or, there might be some prohibited region of the x, y-plane through which the path is disallowed from passing. And so on.

In summary, in the simplest problem in the calculus of variations we are required to find a function $\phi(x) \in C^1[0,1]$ that minimizes a functional $F\{\phi\}$ of the form

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi')dx$$

over an admissible set of test functions A. The test functions (or admissible functions) ϕ are subject to certain conditions including smoothness requirements; possibly (but not necessarily) boundary conditions at both ends $x = 0, 1$; and possibly (but not necessarily) side constraints of various forms. Other types of problems will be also be encountered in what follows.

7.2 Brief review of calculus.

Perhaps it is useful to begin by reviewing the familiar question of minimization in calculus. Consider a subset A of n-dimensional space $\mathbb{R}^n$ and let $F(\mathbf{x}) = F(x_1, x_2, \ldots, x_n)$ be a real-valued function defined on A. We say that $\mathbf{x}_o \in \mathsf{A}$ is a *minimizer* of F if[1]

$$F(\mathbf{x}) \geq F(\mathbf{x}_o) \qquad \text{for all} \quad \mathbf{x} \in \mathsf{A}. \tag{7.3}$$

Sometimes we are only interested in finding a "local minimizer", i.e. a point $\mathbf{x}_o$ that minimizes F relative to all $\mathbf{x}$ that are "close" to $\mathbf{x}_0$. In order to speak of such a notion we must have a measure of "closeness". Thus suppose that the vector space $\mathbb{R}^n$ is Euclidean so that a norm is defined on $\mathbb{R}^n$. Then we say that $\mathbf{x}_o$ is a *local minimizer* of F if $F(\mathbf{x}) \geq F(\mathbf{x}_o)$ for all $\mathbf{x}$ in a neighborhood of $\mathbf{x}_o$, i.e. if

$$F(\mathbf{x}) \geq F(\mathbf{x}_o) \quad \text{for all} \quad \mathbf{x} \quad \text{such that} \quad |\mathbf{x} - \mathbf{x}_o| < r \tag{7.4}$$

for some $r > 0$.

Define the function $\hat{F}(\varepsilon)$ for $-\varepsilon_0 < \varepsilon < \varepsilon_0$ by

$$\hat{F}(\varepsilon) = F(\mathbf{x}_0 + \varepsilon\mathbf{n}) \tag{7.5}$$

where $\mathbf{n}$ is a fixed vector and ε_0 is small enough to ensure that $\mathbf{x}_0 + \varepsilon\mathbf{n} \in \mathsf{A}$ for all $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$. In the presence of sufficient smoothness we can write

$$\hat{F}(\varepsilon) - \hat{F}(0) = \hat{F}'(0)\varepsilon + \frac{1}{2}\hat{F}''(0)\varepsilon^2 + O(\varepsilon^3). \tag{7.6}$$

Since $F(\mathbf{x}_0 + \varepsilon\mathbf{n}) \geq F(\mathbf{x}_0)$ it follows that $\hat{F}(\varepsilon) \geq \hat{F}(0)$. Thus if $\mathbf{x}_0$ is to be a minimizier of F it is necessary that

$$\hat{F}'(0) = 0, \qquad \hat{F}''(0) \geq 0. \tag{7.7}$$

[1] A maximizer of F is a minimizer of $-F$ so we don't need to address maximizing separately from minimizing.

It is customary to use the following notation and terminology: we set

$$\delta F(\mathbf{x}_o, \mathbf{n}) = \hat{F}'(0), \tag{7.8}$$

which is called the *first variation* of F and similarly set

$$\delta^2 F(\mathbf{x}_o, \mathbf{n}) = \hat{F}''(0) \tag{7.9}$$

which is called the *second variation* of F. At an interior local minimizer $\mathbf{x}_0$, one necessarily must have

$$\delta F(\mathbf{x}_o, \mathbf{n}) = 0 \qquad \text{and} \qquad \delta^2 \mathrm{F}(\mathbf{x}_\mathrm{o}, \mathbf{n}) \geq 0 \quad \text{for all unit vectors } \mathbf{n}. \tag{7.10}$$

In the present setting of calculus, we know from (7.5), (7.8) that $\delta F(\mathbf{x}_o, \mathbf{n}) = \boldsymbol{\nabla} F(\mathbf{x}_o) \cdot \mathbf{n}$ and that $\delta^2 F(\mathbf{x}_o, \mathbf{n}) = (\boldsymbol{\nabla}\boldsymbol{\nabla} F(\mathbf{x}_o))\mathbf{n}\cdot\mathbf{n}$. Here the vector field $\boldsymbol{\nabla} F$ is the gradient of F and the tensor field $\boldsymbol{\nabla}\boldsymbol{\nabla} F$ is the gradient of $\boldsymbol{\nabla} F$. Therefore (7.10) is equivalent to the requirements that

$$\boldsymbol{\nabla} F(\mathbf{x}_o) \cdot \mathbf{n} = 0 \qquad \text{and} \qquad (\boldsymbol{\nabla}\boldsymbol{\nabla}\mathrm{F}(\mathbf{x}_\mathrm{o}))\mathbf{n} \cdot \mathbf{n} \geq 0 \quad \text{for all unit vectors } \mathbf{n} \tag{7.11}$$

or equivalently

$$\sum_{i=1}^{n} \left.\frac{\partial F}{\partial x_i}\right|_{\mathbf{x}=\mathbf{x}_0} n_i = 0 \qquad \text{and} \qquad \sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}} \left.\frac{\partial^2 \mathrm{F}}{\partial \mathrm{x_i} \partial \mathrm{x_j}}\right|_{\mathbf{x}=\mathbf{x}_0} \mathrm{n_i n_j} \geq 0 \tag{7.12}$$

whence we must have $\boldsymbol{\nabla} F(\mathbf{x}_o) = \mathbf{o}$ and the Hessian $\boldsymbol{\nabla}\boldsymbol{\nabla} F(\mathbf{x}_o)$ must be positive semi-definite.

Remark: It is worth recalling that a function need not have a minimizer. For example, the function $F_1(x) = x$ defined on $\mathsf{A}_1 = (-\infty, \infty)$ is unbounded as $x \to \pm\infty$. Another example is given by the function $F_2(x) = x$ defined on $\mathsf{A}_2 = (-1, 1)$ noting that $F_2 \geq -1$ on A_2; however, while the value of F_2 can get as close as one wishes to -1, it cannot actually achieve the value -1 since there is no $x \in \mathsf{A}_2$ at which $f(x) = -1$; note that $-1 \notin \mathsf{A}_2$. Finally, consider the function $F_3(x)$ defined on $\mathsf{A}_3 = [-1, 1]$ where $F_3(x) = 1$ for $-1 \leq x \leq 0$ and $F(x) = x$ for $0 < x \leq 1$; the value of F_3 can get as close as one wishes to 0 but cannot achieve it since $F(0) = 1$. In the first example A_1 was unbounded. In the second, A_2 was bounded but open. And in the third example A_3 was bounded and closed but the function was discontinuous on A_3. In order for a minimizer to exist, A must be compact (i.e. bounded and closed). It can be shown that if A is compact and if F is continuous on A then F assumes both maximum and minimum values on A.

7.3 The basic idea: necessary conditions for a minimum: $\delta F = 0$, $\delta^2 F \geq 0$.

In the calculus of variations, we are typically given a functional F defined on a function space A, where $F : \mathsf{A} \to \mathbb{R}$, and we are asked to find a function $\phi_o \in \mathsf{A}$ that minimizes F over A: i.e. to find $\phi_o \in \mathsf{A}$ for which

$$F\{\phi\} \geq F\{\phi_o\} \qquad \text{for all} \quad \phi \in \mathsf{A}.$$

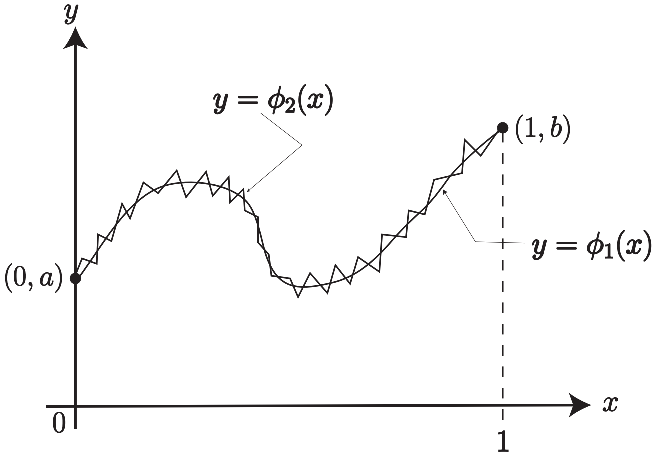


Figure 7.2: Two functions ϕ_1 and ϕ_2 that are "close" in the sense of the norm $||\cdot||_0$ but not in the sense of the norm $||\cdot||_1$.

Most often, we will be looking for a *local (or relative) minimizer*, i.e. for a function ϕ_0 that minimizes F relative to all "nearby functions". This requires that we select a norm so that the distance between two functions can be quantified. For a function ϕ in the set of functions that are continuous on an interval $[x_1, x_2]$, i.e. for $\phi \in C[x_1, x_2]$, one can define a norm by

$$||\phi||_0 = \max_{x_1 \leq x \leq x_2} |\phi(x)|.$$

For a function ϕ in the set of functions that are continuous and have continuous first derivatives on $[x_1, x_2]$, i.e. for $\phi \in C^1[x_1, x_2]$ one can define a norm by

$$||\phi||_1 = \max_{x_1 \leq x \leq x_2} |\phi(x)| \;+\; \max_{x_1 \leq x \leq x_2} |\phi'(x)|;$$

and so on. (Of course the norm $||\phi||_0$ can also be used on $C^1[x_1, x_2]$.)

When seeking a local minimizer of a functional F we might say we want to find ϕ_0 for which

$$F\{\phi\} \geq F\{\phi_o\} \qquad \text{for all admissible } \phi \text{ such that} \quad ||\phi - \phi_0||_0 < r$$

for some $r > 0$. In this case the minimizer ϕ_0 is being compared with all admissible functions ϕ whose values are close to those of ϕ_0 for all $x_1 \leq x \leq x_2$. Such a local minimizer is called a *strong minimizer*. On the other hand, when seeking a local minimizer we might say we want to find ϕ_0 for which

$$F\{\phi\} \geq F\{\phi_o\} \qquad \text{for all admissible } \phi \text{ such that} \quad ||\phi - \phi_0||_1 < r$$

for some $r > 0$. In this case the minimizer is being compared with all functions whose values *and whose first derivatives* are close to those of ϕ_0 for all $x_1 \leq x \leq x_2$. Such a local minimizer is called a *weak minimizer*. A strong minimizer is automatically a weak minimizer.

Unless explicitly stated otherwise, in this Chapter we will be examining weak local extrema. The approach for finding such extrema of a functional is essentially the same as that used in the more familiar case of calculus reviewed in the preceding sub-section. Consider a functional $F\{\phi\}$ defined on a function space A and suppose that $\phi_o \in \mathsf{A}$ minimizes F. In order to determine ϕ_0 we consider the one-parameter family of admissible functions

$$\phi(x;\varepsilon) = \phi_0(x) + \varepsilon\,\eta(x) \tag{7.13}$$

that are close to ϕ_0; here ε is a real variable in the range $-\varepsilon_0 < \varepsilon < \varepsilon_0$ and $\eta(x)$ is a once continuously differentiable function. Since ϕ is to be admissible, we must have $\phi_0 + \varepsilon\eta \in \mathsf{A}$ for each $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$. Define a function $\hat{F}(\varepsilon)$ by

$$\hat{F}(\varepsilon) = F\{\phi_0 + \varepsilon\eta\}, \qquad -\varepsilon_0 < \varepsilon < \varepsilon_0. \tag{7.14}$$

Since ϕ_0 minimizes F it follows that $F\{\phi_0 + \varepsilon\eta\} \geq F\{\phi_0\}$ or equivalently $\hat{F}(\varepsilon) \geq \hat{F}(0)$. Therefore $\varepsilon = 0$ minimizes $\hat{F}(\varepsilon)$. The first and second variations of F are defined by $\delta F\{\phi_0, \eta\} = \hat{F}'(0)$ and $\delta^2 F\{\phi_0, \eta\} = \hat{F}''(0)$ respectively, and so if ϕ_0 minimizes F, then it is necessary that

$$\delta F\{\phi_0, \eta\} = 0, \qquad \delta^2 F\{\phi_0, \eta\} \geq 0. \tag{7.15}$$

These are necessary conditions on a minimizer ϕ_o. We cannot go further in general. In any specific problem, such as those in the subsequent sections, the necessary condition $\delta F\{\phi_o, \eta\} = 0$ can be further simplified by exploiting the fact that it must hold for all admissible η. This allows one to eliminate η leading to a condition (or conditions) that only involves the minimizer ϕ_0.

Remark: Note that when η is independent of ε the functions $\phi_0(x)$ and $\phi_0(x)+\varepsilon\eta(x)$, and their derivatives, are close to each other for small ε. On the other hand the functions $\phi_0(x)$ and $\phi_0(x)+\varepsilon\sin(x/\varepsilon)$ are close to each other but their derivatives are not close to each other. Throughout these notes we will consider functions η that are independent of ε and so, as noted previously, we will be restricting attention exclusively to weak minimizers.

7.4 Application of the necessary condition $\delta F = 0$ to the basic problem. Euler equation.

7.4.1 The basic problem. Euler equation.

Consider the following class of problems: let A be the set of all continuously differentiable functions $\phi(x)$ defined for $0 \le x \le 1$ with $\phi(0)=a$, $\phi(1)=b$:

$$\mathsf{A} = \big\{\phi(\cdot)\,\big|\; \phi:[0,1]\to\mathbb{R},\ \phi\in C^1[0,1],\ \phi(0)=a,\ \phi(1)=b\big\}\,. \tag{7.16}$$

Let $f(x,y,z)$ be a given function, defined and smooth for all real x,y,z. Define a *functional* $F\{\phi\}$, for every $\phi\in\mathsf{A}$, by

$$F\{\phi\} = \int_0^1 f\left(x,\phi(x),\phi'(x)\right)\,dx. \tag{7.17}$$

We wish to find a function $\phi\in\mathsf{A}$ which minimizes $F\{\phi\}$.

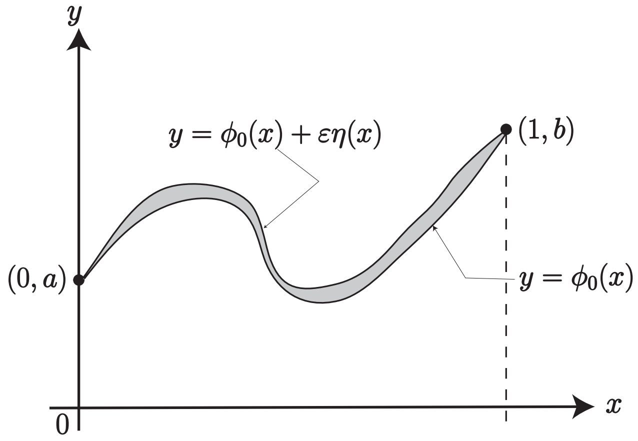


Figure 7.3: The minimizer ϕ_0 and a neighboring function $\phi_0+\varepsilon\eta$.

Suppose that $\phi_0(x) \in \mathsf{A}$ is a minimizer of F, so that $F\{\phi\} \geq F\{\phi_0\}$ for all $\phi \in \mathsf{A}$. In order to determine ϕ_0 we consider the one parameter family of admissible functions $\phi(x;\varepsilon) = \phi_0(x) + \varepsilon\,\eta(x)$ where ε is a real variable in the range $-\varepsilon_0 < \varepsilon < \varepsilon_0$ and $\eta(x)$ is a once continuously differentiable function on $[0,1]$; see Figure 7.3. Since ϕ must be admissible we need $\phi_0 + \varepsilon\,\eta \in \mathsf{A}$ for each ε. Therefore we must have $\phi(0,\varepsilon) = a$ and $\phi(1,\varepsilon) = b$ which in turn requires that

$$\eta(0) = \eta(1) = 0. \tag{7.18}$$

Pick any function $\eta(x)$ with the property (7.18) and fix it. Define the function $\hat{F}(\varepsilon) = F\{\phi_0 + \varepsilon\eta\}$ so that

$$\hat{F}(\varepsilon) = F\{\phi_0 + \varepsilon\eta\} = \int_0^1 f(x,\ \phi_0 + \varepsilon\eta,\ \phi_0' + \varepsilon\eta')\,dx. \tag{7.19}$$

We know from the analysis of the preceding section that a necessary condition for ϕ_0 to minimize F is that

$$\delta F\{\phi_o, \eta\} = \hat{F}'(0) = 0. \tag{7.20}$$

On using the chain-rule, we find $\hat{F}'(\varepsilon)$ from (7.19) to be

$$\hat{F}'(\varepsilon) = \int_0^1 \left(\frac{\partial f}{\partial y}(x,\ \phi_0 + \varepsilon\eta,\ \phi_0' + \varepsilon\eta')\,\eta + \frac{\partial f}{\partial z}(x,\ \phi_0 + \varepsilon\eta,\ \phi_0' + \varepsilon\eta')\eta' \right)\,dx,$$

and so (7.20) leads to

$$\delta F\{\phi_o, \eta\} = \hat{F}'(0) = \int_0^1 \left(\frac{\partial f}{\partial y}(x, \phi_0, \phi_0')\eta + \frac{\partial f}{\partial z}(x, \phi_0, \phi_0')\eta' \right)\,dx \ = \ 0. \tag{7.21}$$

Thus far we have simply repeated the general analysis of the preceding section in the context of the particular functional (7.17). Our goal is to find ϕ_0 and so we must eliminate η from (7.21). To do this we rearrange the terms in (7.21) into a convenient form and exploit the fact that (7.21) must hold for all functions η that satisfy (7.18).

In order to do this we proceed as follows: Integrating the second term in (7.21) by parts gives

$$\int_0^1 \left(\frac{\partial f}{\partial z}\right)\eta'\,dx = \left[\eta\,\frac{\partial f}{\partial z}\right]_{x=0}^{x=1} - \int_0^1 \frac{d}{dx}\left(\frac{\partial f}{\partial z}\right)\,\eta\,dx\ .$$

However by (7.18) we have $\eta(0) = \eta(1) = 0$ and therefore the first term on the right-hand side drops out. Thus (7.21) reduces to

$$\int_0^1 \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial z}\right)\right]\,\eta\,dx = 0. \tag{7.22}$$

Though we have viewed η as fixed up to this point, we recognize that the above derivation is valid for *all* once continuously differentiable functions $\eta(x)$ which have $\eta(0) = \eta(1) = 0$. Therefore (7.22) must hold for *all* such functions.

Lemma: The following is a basic result from calculus: Let $p(x)$ be a *continuous* function on $[0, 1]$ and suppose that

$$\int_0^1 p(x)n(x)dx \ = \ 0$$

for *all* continuous functions $n(x)$ with $n(0) = n(1) = 0$. Then,

$$p(x) = 0 \qquad \text{for} \quad 0 \leq x \leq 1.$$

In view of this Lemma we conclude that the integrand of (7.22) must vanish and therefore obtain the differential equation

$$\frac{d}{dx}\left[\frac{\partial f}{\partial z}(x, \phi_0, \phi_0')\right] - \frac{\partial f}{\partial y}(x, \phi_0, \phi_0') = 0 \quad \text{for} \quad 0 \leq x \leq 1. \tag{7.23}$$

This is a differential equation for ϕ_0, which together with the boundary conditions

$$\phi_0(0) = a, \quad \phi_0(1) = b, \tag{7.24}$$

provides the mathematical problem governing the minimizer $\phi_0(x)$. The differential equation (7.23) is referred to as the *Euler equation* (sometimes referred to as the Euler-Lagrange equation) associated with the functional (7.17).

Notation: In order to avoid the (precise though) cumbersome notation above, we shall drop the subscript "0" from the minimizing function ϕ_0; moreover, we shall write the Euler equation (7.23) as

$$\frac{d}{dx}\left[\frac{\partial f}{\partial \phi'}(x, \phi, \phi')\right] - \frac{\partial f}{\partial \phi}(x, \phi, \phi') = 0, \tag{7.25}$$

where, in carrying out the *partial* differentiation in (7.25), one treats x, ϕ and ϕ' as if they were independent variables.

7.4.2 An example. The Brachistochrone Problem.

Consider the Brachistochrone Problem formulated in the first example of Section 7.1. Here we have

$$f(x, \phi, \phi') = \sqrt{\frac{1 + (\phi')^2}{2g\phi}}$$

and we wish to find the function $\phi_0(x)$ that minimizes

$$F\{\phi\} = \int_0^1 f(x, \phi(x), \phi'(x))dx \ = \ \int_0^1 \sqrt{\frac{1 + [\phi'(x)]^2}{2g\phi(x)}}\, dx$$

over the class of functions $\phi(x)$ that are continuous and have continuous first derivatives on [0,1], and satisfy the boundary conditions $\phi(0) = 0, \ \phi(1) = h$. Treating x, ϕ and ϕ' as if they are independent variables and differentiating the function $f(x, \phi, \phi')$ gives:

$$\frac{\partial f}{\partial \phi} = \sqrt{\frac{1 + (\phi')^2}{2g}}\ \frac{1}{2(\phi)^{3/2}}, \qquad\qquad \frac{\partial f}{\partial \phi'} = \frac{\phi'}{\sqrt{2g\phi(1 + (\phi')^2)}}\ ,$$

and therefore the Euler equation (7.23) specializes to

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\phi'}{\sqrt{(\phi)(1 + (\phi')^2)}}\right) \ - \ \frac{\sqrt{1 + (\phi')^2}}{2(\phi)^{3/2}} \ = \ 0, \qquad 0 < x < 1, \tag{7.26}$$

with associated boundary conditions

$$\phi(0) = 0, \qquad \phi(1) = h. \tag{7.27}$$

The minimizer $\phi(x)$ therefore must satisfy the boundary-value problem consisting of the second-order (nonlinear) ordinary differential equation (7.26) and the boundary conditions (7.27).

The rest of this sub-section has nothing to do with the calculus of variations. It is simply concerned with the solving the boundary value problem (7.26), (7.27). We can write the differential equation as

$$\frac{\phi'}{\sqrt{\phi(1 + (\phi')^2)}}\, \frac{\mathrm{d}}{\mathrm{d}\phi}\left(\frac{\phi'}{\sqrt{\phi(1 + (\phi')^2)}}\right) \ + \ \frac{1}{2\phi^2} \ = \ 0$$

which can be immediately integrated to give

$$\frac{1}{(\phi'(x))^2} = \frac{\phi(x)}{c^2 - \phi(x)} \tag{7.28}$$

where c is a constant of integration that is to be determined.

It is most convenient to find the path of fastest descent in parametric form, $x = x(\theta), \phi = \phi(\theta), \theta_1 < \theta < \theta_2$, and to this end we adopt the substitution

$$\phi = \frac{c^2}{2}(1 - \cos\theta) \ = \ c^2 \sin^2(\theta/2), \qquad \theta_1 < \theta < \theta_2. \tag{7.29}$$

Differentiating this with respect to x gives

$$\phi'(x) = \frac{c^2}{2}\sin\theta\ \theta'(x)$$

so that, together with (7.28) and (7.29), this leads to

$$\frac{dx}{d\theta} = \frac{c^2}{2}(1-\cos\theta)$$

which integrates to give

$$x = \frac{c^2}{2}(\theta - \sin\theta) + c_1, \qquad \theta_1 < \theta < \theta_2. \tag{7.30}$$

We now turn to the boundary conditions. The requirement $\phi(x) = 0$ at $x = 0$, together with (7.29) and (7.30), gives us $\theta_1 = 0$ and $c_1 = 0$. We thus have

$$\left.\begin{aligned} x &= \frac{c^2}{2}(\theta - \sin\theta), \\ \phi &= \frac{c^2}{2}(1-\cos\theta), \end{aligned}\right\} \qquad 0 \le \theta \le \theta_2. \tag{7.31}$$

The remaining boundary condition $\phi(x) = h$ at $x = 1$ gives the following two equations for finding the two constants θ_2 and c:

$$\left.\begin{aligned} 1 &= \frac{c^2}{2}(\theta_2 - \sin\theta_2), \\ h &= \frac{c^2}{2}(1-\cos\theta_2). \end{aligned}\right\} \tag{7.32}$$

Once this pair of equations is solved for c and θ_2 then (7.31) provides the solution of the problem. We now address the solvability of (7.32).

To this end, first, if we define the function $p(\theta)$ by

$$p(\theta) = \frac{\theta - \sin\theta}{1-\cos\theta}, \qquad 0 < \theta < 2\pi, \tag{7.33}$$

then, by dividing the first equation in (7.32) by the second, we see that θ_2 is a root of the equation

$$p(\theta_2) = h^{-1}. \tag{7.34}$$

One can readily verify that the function $p(\theta)$ has the properties

$$p \to 0 \quad \text{as } \theta \to 0+, \qquad p \to \infty \quad \text{as } \theta \to 2\pi-,$$

$$\frac{dp}{d\theta} = \frac{\cos\theta/2}{\sin^3\theta/2}\left(\tan\theta/2 - \theta/2\right) > 0 \quad \text{for } 0 < \theta < 2\pi.$$

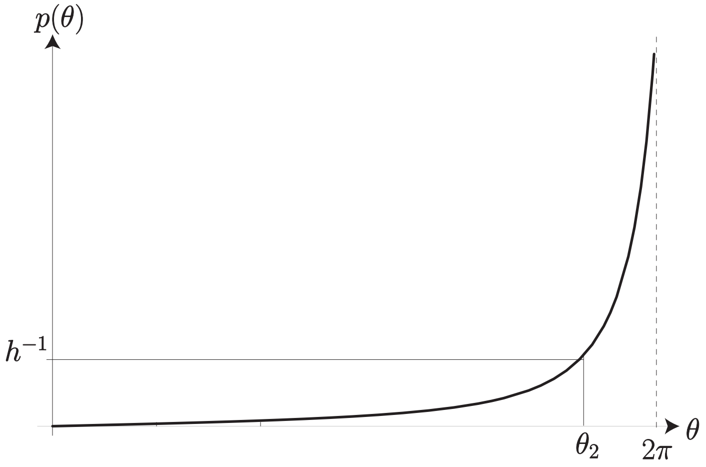
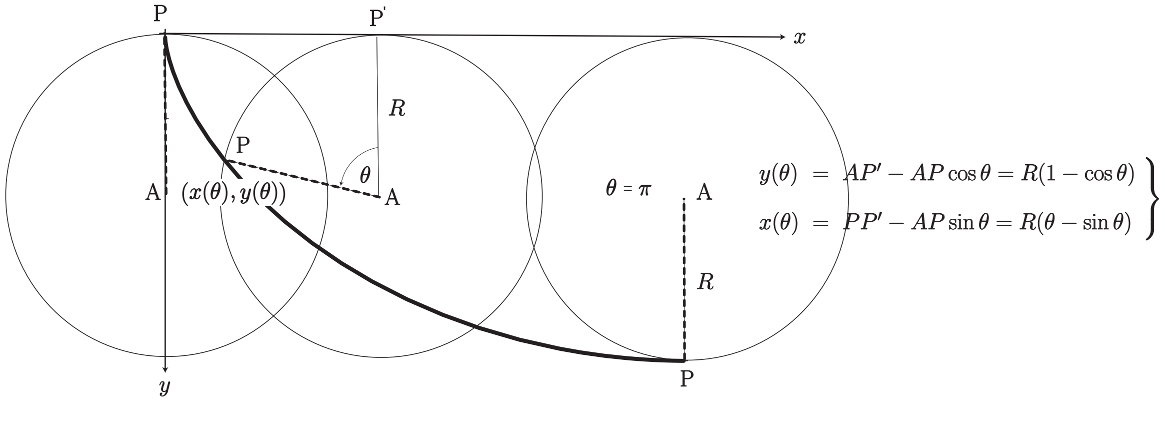


Figure 7.4: A graph of the function $p(\theta)$ defined in (7.33) versus θ. Note that given any $h > 0$ the equation $h^{-1} = p(\theta)$ has a unique root $\theta = \theta_2 \in (0, 2\pi)$.

Therefore it follows that as θ goes from 0 to 2π, the function $p(\theta)$ increases monotonically from 0 to ∞; see Figure 7.4. Therefore, given any $h > 0$, the equation $p(\theta_2) = h^{-1}$ can be solved for a unique value of $\theta_2 \in (0, 2\pi)$. The value of c is then given by $(7.32)_1$.

Thus in summary, the path of minimum descent is given by the curve defined in (7.31) with the values of θ_2 and c given by (7.34) and $(7.32)_1$ respectively. Figure 7.5 shows that the curve (7.31) is a cycloid – the path traversed by a point on the rim of a wheel that rolls without slipping.



P
P'
x
R
P
θ
A
$(x(\theta), y(\theta))$
A
$\theta = \pi$
A
R
P
y
$y(\theta) = AP' - AP\cos\theta = R(1 - \cos\theta)$
$x(\theta) = PP' - AP\sin\theta = R(\theta - \sin\theta)$

Figure 7.5: A cycloid $x = x(\theta), y = y(\theta)$ is generated by rolling a circle along the x-axis as shown, the parameter θ having the significance of being the angle of rolling.

7.4.3 A Formalism for Deriving the Euler Equation

In order to expedite the steps involved in deriving the Euler equation, one usually uses the following formal procedure. First, we adopt the following *notation*: if $H\{\phi\}$ is any quantity that depends on ϕ, then by δH we mean[2]

$$\delta H = H(\phi + \varepsilon\eta) - H(\phi) \quad \text{up to linear terms.} \tag{7.35}$$

that is,

$$\delta H = \varepsilon \left. \frac{dH\{\phi+\varepsilon\eta\}}{d\varepsilon} \right|_{\varepsilon=0}, \tag{7.36}$$

For example, by $\delta\phi$ we mean

$$\delta\phi = (\phi + \varepsilon\eta) - (\phi) = \varepsilon\eta; \tag{7.37}$$

by $\delta\phi'$ we mean

$$\delta\phi' = (\phi' + \varepsilon\eta') - (\phi') = \varepsilon\eta' = (\delta\phi)'; \tag{7.38}$$

by δf we mean

$$\begin{aligned} \delta f &= f(x,\ \phi+\varepsilon\eta,\ \phi'+\varepsilon\eta') - f(x,\phi,\phi') \\ &= \frac{\partial f}{\partial\phi}(x,\phi,\phi')\,\varepsilon\eta + \frac{\partial f}{\partial\phi'}(x,\phi,\phi')\,\varepsilon\eta' \\ &= \left(\frac{\partial f}{\partial\phi}\right)\,\delta\phi + \left(\frac{\partial f}{\partial\phi'}\right)\,\delta\phi'; \end{aligned} \tag{7.39}$$

and by δF, or $\delta \int_0^1 f\,dx$, we mean

$$\begin{aligned} \delta F &= F\{\phi+\varepsilon\eta\} - F\{\phi\} = \varepsilon\left[\frac{d}{d\varepsilon}F\{\phi+\varepsilon\eta\}\right]_{\varepsilon=0} \\ &= \varepsilon\int_0^1 \left[\left(\frac{\partial f}{\partial\phi}\right)\eta + \left(\frac{\partial f}{\partial\phi'}\right)\eta'\right]\,dx \\ &= \int_0^1 \left[\frac{\partial f}{\partial\phi}\,\delta\phi + \frac{\partial f}{\partial\phi'}\,\delta\phi'\right]\,dx = \int_0^1 \delta f\,dx. \end{aligned} \tag{7.40}$$

We refer to $\delta\phi(x)$ as an *admissible variation*. When $\eta(0) = \eta(1) = 0$, it follows that

$$\delta\phi(0) = \delta\phi(1) = 0.$$

[2] Note the following minor change in notation: what we call δH here is what we previously would have called $\varepsilon\,\delta H$.

We refer to δF as the *first variation of the functional* F. Observe from (7.40) that

$$\delta F = \delta \int_0^1 f\, dx = \int_0^1 \delta f\, dx. \tag{7.41}$$

Finally observe that the necessary condition for a minimum that we wrote down previously can be written as

$$\delta F\{\phi, \delta\phi\} = 0 \qquad \text{for all admissible variations} \quad \delta\phi. \tag{7.42}$$

For purposes of illustration, let us now repeat our previous derivation of the Euler equation using this new notation[3]. Given the functional F, a necessary condition for an extremum of F is

$$\delta F = 0$$

and so our task is to calculate δF:

$$\delta F = \delta \int_0^1 f\, dx = \int_0^1 \delta f\, dx.$$

Since $f = f(x, \phi, \phi')$, this in turn leads to[4]

$$\delta F = \int_0^1 \left[\left(\frac{\partial f}{\partial \phi} \right) \delta\phi + \left(\frac{\partial f}{\partial \phi'} \right) \delta\phi' \right] dx.$$

From here on we can proceed as before by setting $\delta F = 0$, integrating the second term by parts, and using the boundary conditions and the arbitrariness of an admissible variation $\delta\phi(x)$ to derive the Euler equation.

7.5 Generalizations.

7.5.1 Generalization: Free end-point; Natural boundary conditions.

Consider the following modified problem: suppose that we want to find the function $\phi(x)$ from among all once continuously differentiable functions that makes the functional

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi')\, dx$$

[3] If ever in doubt about a particular step during a calculation, always go back to the meaning of the symbols $\delta\phi$, etc. or revert to using ε, η etc.

[4] Note that the variation δ does not operate on x since it is the function ϕ that is being varied not the independent variable x. So in particular, $\delta f = f_\phi \delta\phi + f_{\phi'}\delta\phi'$ and *not* $\delta f = f_x \delta x + f_\phi \delta\phi + f_{\phi'}\delta\phi'$.

a minimum. Note that we do *not* restrict attention here to those functions that satisfy $\phi(0) = a$, $\phi(1) = b$. So the set of admissible functions A is

$$\mathsf{A} = \big\{\phi(\cdot)|\ \phi : [0,1] \to \mathbb{R},\ \phi \in C^1[0,1]\big\} \tag{7.43}$$

Note that the class of admissible functions A is much larger than before. The functional $F\{\phi\}$ is defined for all $\phi \in \mathsf{A}$ by

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi')\, dx. \tag{7.44}$$

We begin by calculating the first variation of F:

$$\delta F = \delta \int_0^1 f\, dx = \int_0^1 \delta f\, dx = \int_0^1 \left[\left(\frac{\partial f}{\partial \phi}\right)\delta\phi + \left(\frac{\partial f}{\partial \phi'}\right)\delta\phi'\right]\, dx \tag{7.45}$$

Integrating the last term by parts yields

$$\delta F = \int_0^1 \left[\frac{\partial f}{\partial \phi} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\right]\delta\phi\, dx + \left[\frac{\partial f}{\partial \phi'}\delta\phi\right]_0^1. \tag{7.46}$$

Since $\delta F = 0$ at an extremum, we must have

$$\int_0^1 \left[\frac{\partial f}{\partial \phi} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\right]\delta\phi\, dx + \left[\frac{\partial f}{\partial \phi'}\delta\phi\right]_0^1 \ = \ 0 \tag{7.47}$$

for all admissible variations $\delta\phi(x)$. Note that the boundary term in (7.47) does not automatically drop out now because $\delta\phi(0)$ and $\delta\phi(1)$ do not have to vanish. First restrict attention to all variations $\delta\phi$ with the additional property $\delta\phi(0) = \delta\phi(1) = 0$; equation (7.47) must necessarily hold for all such variations $\delta\phi$. The boundary terms now drop out and by the Lemma in Section 7.4.1 it follows that

$$\frac{d}{dx}\left[\frac{\partial f}{\partial \phi'}\right] - \frac{\partial f}{\partial \phi} = 0 \quad \text{for} \quad 0 < x < 1. \tag{7.48}$$

This is the same Euler equation as before. Next, return to (7.47) and keep (7.48) in mind. We see that we must have

$$\left.\frac{\partial f}{\partial \phi'}\right|_{x=1} \delta\phi(1) \ - \ \left.\frac{\partial f}{\partial \phi'}\right|_{x=0} \delta\phi(0) = 0 \tag{7.49}$$

for *all* admissible variations $\delta\phi$. Since $\delta\phi(0)$ and $\delta\phi(1)$ are both arbitrary (and not necessarily zero), (7.49) requires that

$$\frac{\partial f}{\partial \phi'} = 0 \text{ at } x = 0 \text{ and } x = 1. \tag{7.50}$$

Equation (7.50) provides the boundary conditions to be satisfied by the extremizing function $\phi(x)$. These boundary conditions were determined as part of the extremization; they are referred to as *natural boundary conditions* in contrast to boundary conditions that are given as part of a problem statement.

Example: Reconsider the Brachistochrone Problem analyzed previously but now suppose that we want to find the shape of the wire that commences from $(0,0)$ and ends *somewhere on the vertical* through $x = 1$; see Figure 7.6. The only difference between this and the first Brachistochrone Problem is that here the set of admissible functions is

$$\mathsf{A}_2 = \big\{\phi(\cdot) \,\big|\, \phi : [0,1] \to \mathbb{R},\ \phi \in C^1[0,1],\ \phi(0) = 0\big\}\,;$$

note that there is no restriction on ϕ at $x = 1$. Our task is to minimize the travel time of the bead $T\{\phi\}$ over the set A_2.

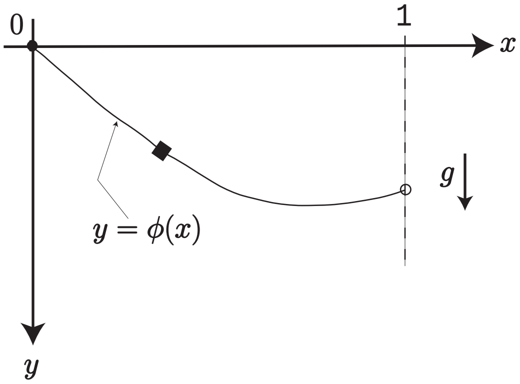


Figure 7.6: Curve joining $(0,0)$ to an arbitrary point on the vertical line through $x = 1$.

The minimizer must satisfy the same Euler equation (7.26) as in the first problem, and the same boundary condition $\phi(0) = 0$ at the left hand end. To find the natural boundary condition at the other end, recall that

$$f(x,\phi,\phi') = \sqrt{\frac{1+(\phi')^2}{2g\phi}}\ .$$

Differentiating this gives

$$\frac{\partial f}{\partial \phi'} = \frac{\phi'}{\sqrt{2g\phi(1+(\phi')^2)}}\ .$$

and so by (7.50), the natural boundary coundition is

$$\frac{\phi'}{\sqrt{2g\phi(1+(\phi')^2)}} = 0 \quad \text{at} \quad x = 1,$$

which simplifies to

$$\phi'(1) = 0.$$

.

7.5.2 Generalization: Higher derivatives.

The functional $F\{\phi\}$ considered above involved a function ϕ and its first derivative ϕ'. One can consider functionals that involve higher derivatives of ϕ, for example

$$F\{\phi\} = \int_0^1 f(x,\, \phi,\, \phi',\, \phi'')\, dx.$$

We begin with the formulation and analysis of a specific example and then turn to some theory.

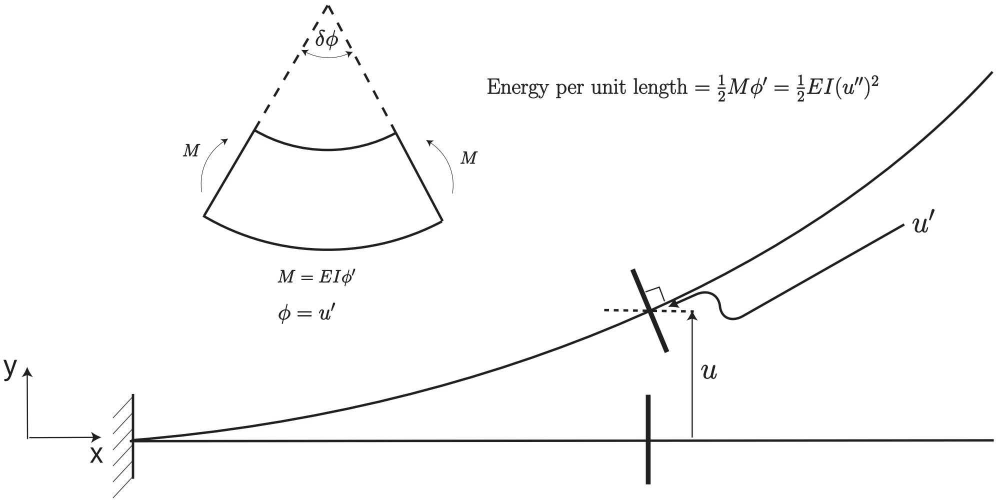

$\delta\phi$
Energy per unit length $= \frac{1}{2}M\phi' = \frac{1}{2}EI(u'')^2$
M
M
u'
$M = EI\phi'$
$\phi = u'$
y
x
u

Figure 7.7: The neutral axis of a beam in reference (straight) and deformed (curved) states. The bold lines represent a cross-section of the beam in the reference and deformed states. In the classical Bernoulli-Euler theory of beams, cross-sections remain perpendicular to the neutral axis.

Example: The Bernoulli-Euler Beam. Consider an elastic beam of length L and bending stiffness EI, which is clamped at its left hand end. The beam carries a distributed load $p(x)$ along its length and a concentrated force F at the right hand end $x = L$; both loads act in the $-y$-direction. Let $u(x)$, $0 \leq x \leq L$, be a geometrically admissible deflection of the beam. Since the beam is clamped at the left hand end this means that $u(x)$ is any (smooth enough) function that satisfies the geometric boundary conditions

$$u(0) = 0, \qquad u'(0) = 0; \tag{7.51}$$

the boundary condition $(7.51)_1$ describes the geometric condition that the beam is clamped at $x = 0$ and therefore cannot deflect at that point; the boundary condition $(7.51)_2$ describes the geometric condition that the beam is clamped at $x = 0$ and therefore cannot rotate at the left end. The set of admissible test functions that we consider is

$$\mathsf{A} = \big\{ u(\cdot) \mid u : [0, L] \to \mathbb{R},\ u \in C^4[0, L],\ u(0) = 0,\ u'(0) = 0 \big\}\ , \tag{7.52}$$

which consists of all "geometrically possible configurations".

From elasticity theory we know that the elastic energy associated with a deformed configuration of the beam is $(1/2)EI(u'')^2$ per unit length. Therefore the total potential energy of the system is

$$\Phi\{u\} = \int_0^L \frac{1}{2} EI(u''(x))^2 \, dx - \int_0^L p(x)u(x) \, dx - F\, u(L), \tag{7.53}$$

where the last two terms represent the potential energy of the distributed and concentrated loading respectively; the negative sign in front of these terms arises because the loads act in the $-y$-direction while u is the deflection in the $+y$-direction. Note that the integrand of the functional involves the higher derivative term u''. In addition, note that only two boundary conditions $u(0) = 0, u'(0) = 0$ are given and so we expect to derive additional natural boundary conditions at the right hand end $x = L$.

The actual deflection of the beam minimizes the potential energy (7.53) over the set (7.52). We proceed in the usual way by calculating the first variation $\delta\Phi$ and setting it equal to zero:

$$\delta\Phi = 0.$$

By using (7.53) this can be written explicitly as

$$\int_0^L EI\, u''\delta u''\, dx - \int_0^L p\ \delta u\, dx - F\ \delta u(L) = 0.$$

Twice integrating the first term by parts leads to

$$\int_0^L EI\, u''''\delta u\, dx - \int_0^L p\ \delta u\, dx - F\ \delta u(L) - \Big[EIu'''\delta u \Big]_0^L + \Big[EIu''\delta u' \Big]_0^L = 0.$$

The given boundary conditions (7.51) require that an admissible variation δu must obey $\delta u(0) = 0, \delta u'(0) = 0$. Therefore the preceding equation simplifies to

$$\int_0^L (EI\, u'''' - p)\, \delta u\, dx - [EIu'''(L) + F]\, \delta u(L) + EIu''(L)\delta u'(L) = 0.$$

Since this must hold for all admissible variations $\delta u(x)$, it follows in the usual way that the extremizing function $u(x)$ must obey

$$\left.\begin{array}{rcll} EI\,u''''(x) - p(x) &=& 0 & \text{for} \quad 0 < x < L, \\[2ex] EI\,u'''(L) + F &=& 0, & \\[2ex] EI\,u''(L) &=& 0. & \end{array}\right\} \tag{7.54}$$

Thus the extremizer $u(x)$ obeys the fourth order linear ordinary differential equation $(7.54)_1$, the prescribed boundary conditions (7.51) and the natural boundary conditions $(7.54)_{2,3}$.

The natural boundary condition $(7.54)_2$ describes the mechanical condition that the beam carries a concentrated force F at the right hand end; and the natural boundary condition $(7.54)_3$ describes the mechanical condition that the beam is free to rotate (and therefore has zero "bending moment") at the right hand end.

Exercise: Consider the functional

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi', \phi'')dx$$

defined on the set of admissible functions A consisting of functions ϕ that are defined and four times continuously differentiable on $[0, 1]$ and that satisfy the four boundary conditions

$$\phi(0) = \phi_0, \quad \phi'(0) = \phi'_0, \quad \phi(1) = \phi_1, \quad \phi'(1) = \phi'_1.$$

Show that the function ϕ that extremizes F over the set A must satisfy the Euler equation

$$\frac{\partial f}{\partial \phi} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial \phi''}\right) = 0 \qquad \text{for} \quad 0 < x < 1$$

where, as before, the partial derivatives $\partial f/\partial\phi$, $\partial f/\partial\phi'$ and $\partial f/\partial\phi''$ are calculated by treating ϕ, ϕ' and ϕ''' as if they are independent variables in $f(x, \phi, \phi', \phi'')$.

7.5.3 Generalization: Multiple functions.

The functional $F\{\phi\}$ considered above involved a single function ϕ and its derivatives. One can consider functionals that involve multiple functions, for example a functional

$$F\{u, v\} = \int_0^1 f(x,\, u,\, u',\, v,\, v')\, dx$$

that involves two functions $u(x)$ and $v(x)$. We begin with the formulation and analysis of a specific example and then turn to some theory.

Example: The Timoshenko Beam. Consider a beam of length L, bending stiffness[5] EI and shear stiffness GA. The beam is clamped at $x = 0$, it carries a distributed load $p(x)$ along its length which acts in the $-y$-direction, and carries a concentrated force F at the end $x = L$, also in the $-y$-direction.

In the simplest model of a beam – the so-called Bernoulli-Euler model – the deformed state of the beam is completely defined by the deflection $u(x)$ of the centerline (the neutral axis) of the beam. In that theory, shear deformations are neglected and therefore a cross-section of the beam remains perpendicular to the neutral axis even in the deformed state. Here we discuss a more general theory of beams, one that accounts for shear deformations.

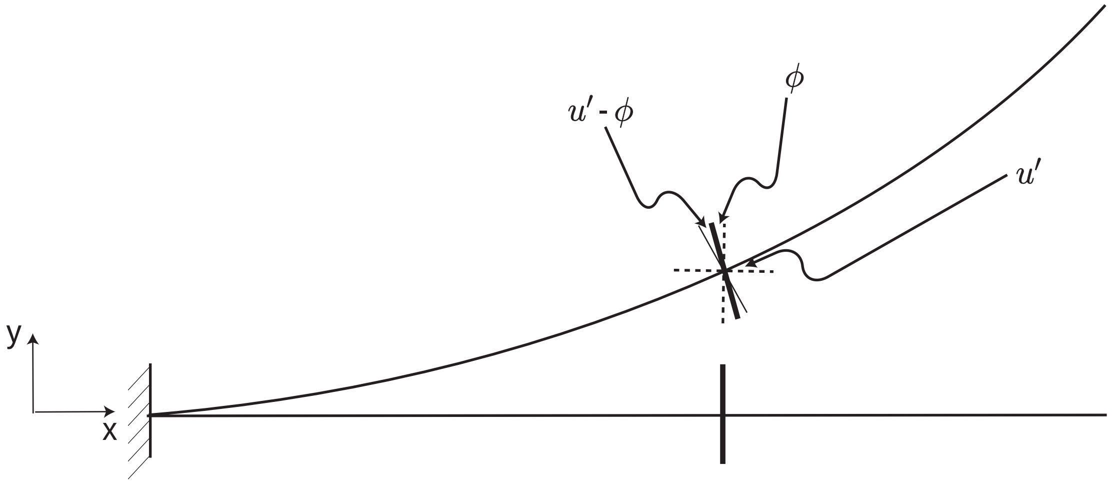


Figure 7.8: The neutral axis of a beam in reference (straight) and deformed (curved) states. The bold lines represent a cross-section of the beam in the reference and deformed states. The thin line is perpendicular to the deformed neutral axis, so that in the classical Bernoulli-Euler theory of beams, where cross-sections remain perpendicular to the neutral axis, the thin line and the bold line would coincide. The angle between the vertical and the bold line if ϕ. The angle between the neutral axis and the horizontal, which equals the angle between the perpendicular to the neutral axis (the thin line) and the vertical dashed line, is u'. The decrease in the angle between the cross-section and the neutral axis is therefore $u' - \phi$.

In the theory considered here, a cross-section of the beam is not constrained to remain perpendicular to the neutral axis. Thus a deformed state of the beam is characterized by *two*

[5] E and G are the Young's modulus and shear modulus of the material, while I and A are the second moment of cross-section and the area of the cross-section respectively.

fields: one, $u(x)$, characterizes the deflection of the centerline of the beam at a location x, and the second, $\phi(x)$, characterizes the rotation of the cross-section at x. (In the Bernoulli-Euler model, $\phi(x) = u'(x)$ since for small angles, the rotation equals the slope.) The fact that the left hand end is clamped implies that the point $x = 0$ cannot deflect and that the cross-section at $x = 0$ cannot rotate. Thus we have the geometric boundary conditions

$$u(0) = 0, \qquad \phi(0) = 0. \tag{7.55}$$

Note that the zero rotation boundary condition is $\phi(0) = 0$ and not $u'(0) = 0$.

In the more accurate beam theory discussed here, the so-called Timoshenko beam theory, one does not neglect shear deformations and so $u(x)$ and $\phi(x)$ are (geometrically) *independent functions*. Since the shear strain is defined as the change in angle between two fibers that are initially at right angles to each other, the shear strain in the present situation is

$$\gamma(x) = u'(x) - \phi(x);$$

see Figure 7.8. Observe that in the Bernoulli-Euler theory $\gamma(x) = 0$.

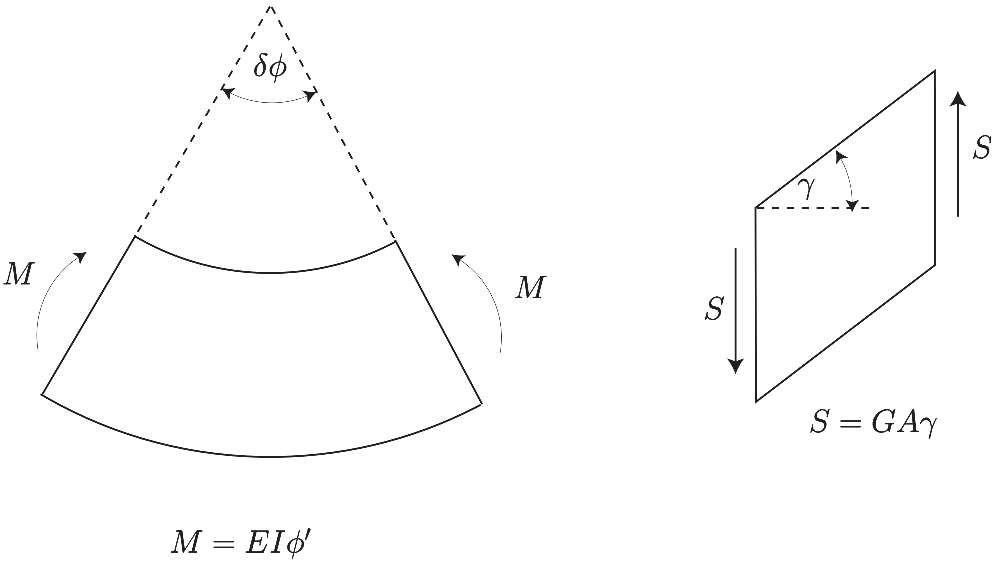


Figure 7.9: Basic constitutive relationships for a beam.

The basic equations of elasticity tell us that the moment-curvature relation for bending is

$$M(x) = EI\phi'(x)$$

and that the associated elastic energy per unit length of the beam, $(1/2)M\phi'$, is

$$\frac{1}{2}EI(\phi'(x))^2.$$

Similarly, we know from elasticity that the shear force-shear strain relation for a beam is[6]

$$S(x) = GA\gamma(x)$$

and that the associated elastic energy per unit length of the beam, $(1/2)S\gamma$, is

$$\frac{1}{2}GA(\gamma(x))^2.$$

The total potential energy of the system is thus

$$\Phi = \Phi\{u,\phi\} = \int_0^L \left\{\frac{1}{2}EI(\phi'(x))^2 \;+\frac{1}{2}GA(u'(x)-\phi(x))^2\right\}\,dx \;-\; \int_0^L pu(x)dx - Fu(L), \tag{7.56}$$

where the last two terms in this expression represent the potential energy of the distributed and concentrated loading respectively (and the negative signs arise because u is the deflection in the $+y$-direction while the loadings p and F are applied in the $-y$-direction). We allow for the possibility that p, EI and GA may vary along the length of the beam and therefore might be functions of x.

The displacement and rotation fields $u(x)$ and $\phi(x)$ associated with an equilibrium configuration of the beam minimizes the potential energy $\Phi\{u,\phi\}$ over the admissible set A of test functions where take

$$\mathsf{A} = \{u(\cdot),\phi(\cdot)\Big|u:[0,l]\to\mathbb{R},\ \phi:[0,l]\to\mathbb{R},\ u\in C^2([0,L]),\ \phi\in C^2([0,L]),\ u(0)=0,\ \phi(0)=0\}.$$

Note that all admissible functions are required to satisfy the geometric boundary conditions (7.55).

To find a minimizer of Φ we calculate its first variation which from (7.56) is

$$\delta\Phi = \int_0^L \Big\{EI\phi'\,\delta\phi' + GA(u'-\phi)(\delta u'-\delta\phi)\Big\}dx - \int_0^L p\,\delta u\,dx - F\,\delta u(L).$$

[6] Since the top and bottom faces of the differential element shown in Figure 7.9 are free of shear traction, we know that the element is not in a state of simple shear. Instead, the shear stress must vary with y such that it vanishes at the top and bottom. In engineering practice, this is taken into account approximately by replacing GA by κGA where the heuristic parameter $\kappa \approx 0.8-0.9$.

Integrating the terms involving $\delta u'$ and $\delta\phi'$ by parts gives

$$\begin{aligned}\delta\Phi &= \Big[EI\phi'\,\delta\phi\Big]_0^L - \int_0^L \frac{d}{dx}\Big(EI\phi'\Big)\delta\phi\,dx \\ &+ \Big[GA(u'-\phi)\delta u\Big]_0^L - \int_0^L \frac{d}{dx}\Big(GA(u'-\phi)\Big)\delta u\,dx - \int_0^L GA(u'-\phi)\delta\phi\,dx \\ &- \int_0^L p\,\delta u\,dx - F\delta u(L).\end{aligned}$$

Finally on using the facts that an admissible variation must satisfy $\delta u(0) = 0$ and $\delta\phi(0) = 0$, and collecting the like terms in the preceding equation leads to

$$\begin{aligned}\delta\Phi &= EI\phi'(L)\,\delta\phi(L) + \Big[GA\Big(u'(L) - \phi(L)\Big) - F\Big]\delta u(L) \\ &- \int_0^L \left[\frac{d}{dx}\Big(EI\phi'\Big) + GA(u'-\phi)\right]\delta\phi(x)\,dx \\ &- \int_0^L \left[\frac{d}{dx}\Big(GA(u'-\phi)\Big) + p\right]\delta u(x)\,dx.\end{aligned} \tag{7.57}$$

At a minimizer, we have $\delta\Phi = 0$ for all admissible variations. Since the variations $\delta u(x)$, $\delta\phi(x)$ are arbitrary on $0 < x < L$ and since $\delta u(L)$ and $\delta\phi(L)$ are also arbitrary, it follows from (7.57) that the field equations

$$\left.\begin{aligned}\frac{d}{dx}\Big(EI\phi'\Big) + GA(u'-\phi) &= 0, \qquad 0 < x < L, \\ \frac{d}{dx}\Big(GA(u'-\phi)\Big) + p &= 0, \qquad 0 < x < L,\end{aligned}\right\} \tag{7.58}$$

and the natural boundary conditions

$$EI\phi'(L) = 0, \qquad GA\Big(u'(L) - \phi(L)\Big) = F \tag{7.59}$$

must hold.

Thus in summary, an equilibrium configuration of the beam is described by the deflection $u(x)$ and rotation $\phi(x)$ that satisfy the differential equations (7.58) and the boundary conditions (7.55), (7.59). [Remark: Can you recover the Bernoulli-Euler theory from the Timoshenko theory in the limit as the shear rigidity $GA \to \infty$?]

Exercise: Consider a smooth function $f(x, y_1, y_2, \ldots, y_n, z_1, z_2, \ldots, z_n)$ defined for all $x, y_1, y_2, \ldots, y_n$, $z_1, \ldots, z_n$. Let $\phi_1(x), \phi_2(x), \ldots, \phi_n(x)$ be n once-continuously differentiable functions on $[0, 1]$

with $\phi_i(0) = a_i, \phi_i(1) = b_i$. Let F be the functional defined by

$$F\{\phi_1, \phi_2, \ldots, \phi_n\} = \int_0^1 f(x, \phi_1, \phi_2, \ldots, \phi_n, \phi_1', \phi_2', \ldots, \phi_n')\, dx \tag{7.60}$$

on the set of all such admissible functions. Show that the functions $\phi_1(x), \phi_2(x), \ldots, \phi_n(x)$ that extremize F must necessarily satisfy the n Euler equations

$$\frac{d}{dx}\left[\frac{\partial f}{\partial \phi_i'}\right] - \frac{\partial f}{\partial \phi_i} = 0 \qquad \text{for} \quad 0 < x < 1, \qquad (i = 1, 2, \ldots, n). \tag{7.61}$$

7.5.4 Generalization: End point of extremal lying on a curve.

Consider the set A of all functions that describe curves in the x, y-plane that commence from a given point $(0, a)$ and end at some point on the curve $G(x, y) = 0$. We wish to minimize a functional $F\{\phi\}$ over this set of functions.

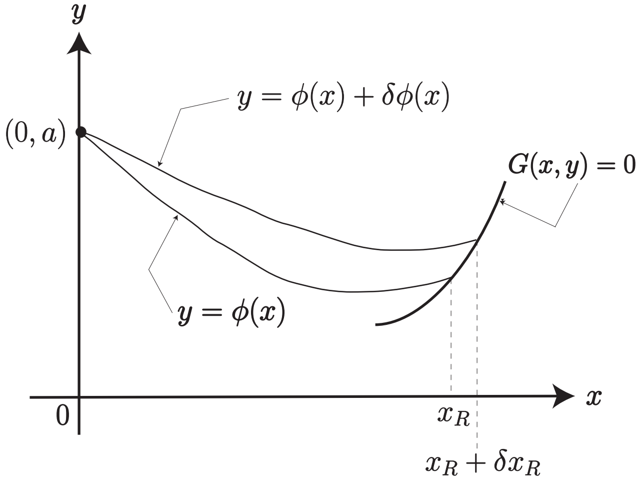


Figure 7.10: Curve joining $(0, a)$ to an arbitrary point on the given curve $G(x, y) = 0$.

Suppose that $\phi(x) \in \mathsf{A}$ is a minimizer of F. Let $x = x_R$ be the abscissa of the point at which the curve $y = \phi(x)$ intersects the curve $G(x, y) = 0$. Observe that x_R is not known a priori and is to be determined along with ϕ. Moreover, note that the abscissa of the point at which a neighboring curve defined by $y = \phi(x) + \delta\phi(x)$ intersects the curve $G = 0$ is not x_R but $x_R + \delta x_R$; see Figure 7.10.

At the minimizer,

$$F\{\phi\} = \int_0^{x_R} f(x, \phi, \phi')dx$$

and at a neighboring test function

$$F\{\phi+\delta\phi\} = \int_0^{x_R+\delta x_R} f(x,\phi+\delta\phi,\phi'+\delta\phi')dx.$$

Therefore on calculating the first variation δF, which equals the linearized form of $F\{\phi+\delta\phi\} - F\{\phi\}$, we find

$$\delta F = \int_0^{x_R+\delta x_R} \Big(f(x,\phi,\phi') + f_\phi(x,\phi,\phi')\delta\phi + f_{\phi'}(x,\phi,\phi')\delta\phi'\Big)dx - \int_0^{x_R} f(x,\phi,\phi')dx$$

where we have set $f_\phi = \partial f/\partial\phi$ and $f_{\phi'} = \partial f/\partial\phi'$. This leads to

$$\delta F = \int_{x_R}^{x_R+\delta x_R} f(x,\phi,\phi')dx \ + \ \int_0^{x_R} \Big(f_\phi(x,\phi,\phi')\delta\phi + f_{\phi'}(x,\phi,\phi')\delta\phi'\Big)dx$$

which in turn reduces to

$$\delta F = f\big(x_R,\phi(x_R),\phi'(x_R)\big)\,\delta x_R \ + \ \int_0^{x_R} \Big(f_\phi\,\delta\phi + f_{\phi'}\,\delta\phi'\Big)dx.$$

Thus setting the first variation δF equal to zero gives

$$f\big(x_R,\phi(x_R),\phi'(x_R)\big)\,\delta x_R \ + \ \int_0^{x_R} \Big(f_\phi\,\delta\phi + f_{\phi'}\,\delta\phi'\Big)dx \ = \ 0.$$

After integrating the last term by parts we get

$$f\big(x_R,\phi(x_R),\phi'(x_R)\big)\delta x_R \ + \ \Big[f_{\phi'}\delta\phi\Big]_0^{x_R} \ + \ \int_0^{x_R} \Big(f_\phi - \frac{d}{dx}f_{\phi'}\Big)\delta\phi\,dx \ = \ 0$$

which, on using the fact that $\delta\phi(0) = 0$, reduces to

$$f\big(x_R,\phi(x_R),\phi'(x_R)\big)\,\delta x_R \ + \ f_{\phi'}\big(x_R,\phi(x_R),\phi'(x_R)\big)\,\delta\phi(x_R) \ + \ \int_0^{x_R} \Big(f_\phi - \frac{d}{dx}f_{\phi'}\Big)\delta\phi\,dx \ = \ 0. \tag{7.62}$$

First limit attention to the subset of all test functions that terminate at the same point $(x_R,\phi(x_R))$ as the minimizer. In this case $\delta x_R = 0$ and $\delta\phi(x_R) = 0$ and so the first two terms in (7.62) vanish. Since this specialized version of equation (7.62) must hold for all such variations $\delta\phi(x)$, this leads to the Euler equation

$$f_\phi - \frac{d}{dx}f_{\phi'} = 0, \qquad 0 \le x \le x_R. \tag{7.63}$$

We now return to arbitrary admissible test functions. Substituting (7.63) into (7.62) gives

$$f\big(x_R,\phi(x_R),\phi'(x_R)\big)\,\delta x_R \ + \ f_{\phi'}\big(x_R,\phi(x_R),\phi'(x_R)\big)\,\delta\phi(x_R) \ = \ 0 \tag{7.64}$$

which must hold for all admissible δx_R and $\delta\phi(x_R)$. It is important to observe that since admissible test curves must end on the curve $G = 0$, the quantities δx_R and $\delta\phi(x_R)$ are *not* independent of each other. Thus (7.64) does *not* hold for *all* δx_R and $\delta\phi(x_R)$; only for those that are consistent with this geometric requirement. The requirement that the minimizing curve and the neighboring test curve terminate on the curve $G(x, y) = 0$ implies that

$$G(x_R, \phi(x_R)) = 0, \qquad G(x_R + \delta x_R, \phi(x_R + \delta x_R) + \delta\phi(x_R + \delta x_R)) = 0, .$$

Note that linearization gives

$$\begin{aligned}
&G(x_R + \delta x_R,\ \phi(x_R + \delta x_R) + \delta\phi(x_R + \delta x_R)) \\
&= G(x_R + \delta x_R,\ \phi(x_R) + \phi'(x_R)\delta x_R + \delta\phi(x_R)) \\
&= G(x_R, \phi(x_R)) + G_x(x_R, \phi(x_R))\delta x_R + G_y(x_R, \phi(x_R))\Big(\phi'(x_R)\delta x_R + \delta\phi(x_R)\Big), \\
&= G(x_R, \phi(x_R)) + \Big(G_x(x_R, \phi(x_R)) + \phi'(x_R)G_y(x_R, \phi(x_R))\Big)\delta x_R + G_y(x_R, \phi(x_R))\, \delta\phi(x_R).
\end{aligned}$$

where we have set $G_x = \partial G/\partial x$ and $G_y = \partial G/\partial x$. Setting $\delta G = G(x_R + \delta x_R, \phi(x_R + \delta x_R) + \delta\phi(x_R + \delta x_R)) - G(x_R, \phi(x_R)) = 0$ thus leads to the following relation between the variations δx_R and $\delta\phi(x_R)$:

$$\Big(G_x(x_R, \phi(x_R)) + \phi'(x_R)G_y(x_R, \phi(x_R))\Big)\delta x_R + G_y(x_R, \phi(x_R))\, \delta\phi(x_R) \;=\; 0. \tag{7.65}$$

Thus (7.64) must hold for all δx_R and $\delta\phi(x_R)$ that satisfy (7.65). This implies that[7]

$$\left.\begin{aligned}
f\big(x_R, \phi(x_R), \phi'(x_R)\big) - \lambda\Big(G_x(x_R, \phi(x_R)) + \phi'(x_R)G_y(x_R, \phi(x_R))\Big) &= 0, \\
f_{\phi'}\big(x_R, \phi(x_R), \phi'(x_R)\big) - \lambda G_y(x_R, \phi(x_R)) &= 0,
\end{aligned}\right\}$$

for some constant λ (referred to as a Lagrange multiplier). We can use the second equation above to simplify the first equation which then leads to the pair of equations

$$\left.\begin{aligned}
f\big(x_R, \phi(x_R), \phi'(x_R)\big) - \phi'(x_R)f_{\phi'}\big(x_R, \phi(x_R), \phi'(x_R)\big) - \lambda G_x(x_R, \phi(x_R)) &= 0, \\
f_{\phi'}\big(x_R, \phi(x_R), \phi'(x_R)\big) - \lambda G_y(x_R, \phi(x_R)) &= 0.
\end{aligned}\right\} \tag{7.66}$$

[7] It may be helpful to recall from calculus that if we are to minimize a function $I(\varepsilon_1, \varepsilon_2)$, we must satisfy the condition $dI = (\partial I/\partial\varepsilon_1)d\varepsilon_1 + (\partial I/\partial\varepsilon_2)d\varepsilon_2 = 0$. But if this minimization is carried out subject to the side constraint $J(\varepsilon_1, \varepsilon_2) = 0$ then we must respect the side condition $dJ = (\partial J/\partial\varepsilon_1)d\varepsilon_1 + (\partial J/\partial\varepsilon_2)d\varepsilon_2 = 0$. Under these circumstances, one finds that that one must require the conditions $\partial I/\partial\varepsilon_1 = \lambda\partial J/\partial\varepsilon_1$, $\partial I/\partial\varepsilon_2 = \lambda\partial J/\partial\varepsilon_2$ where the Lagrange multiplier λ is unknown and is also to be determined. The constrain equation $J = 0$ provides the extra condition required for this purpose.

Equation (7.66) provides two natural boundary conditions at the right hand end $x = x_R$.

In summary: an extremal $\phi(x)$ must satisfy the differential equations (7.63) on $0 \leq x \leq x_R$, the boundary condition $\phi = a$ at $x = 0$, the two natural boundary conditions (7.66) at $x = x_R$, and the equation $G(x_R, \phi(x_R)) = 0$. (Note that the presence of the additional unknown λ is compensated for by the imposition of the additional condition $G(x_R, \phi(x_R)) = 0$.)

Example: Suppose that $G(x, y) = c_1 x + c_2 y + c_3$ and that we are to find the curve of shortest length that commences from $(0, a)$ and ends on $G = 0$.

Since $ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + (\phi')^2}\, dx$ we are to minimize the functional

$$F = \int_0^{x_R} \sqrt{1 + (\phi')^2}\, dx.$$

Thus

$$f(x, \phi, \phi') = \sqrt{1 + (\phi')^2}, \quad f_\phi(x, \phi, \phi') = 0 \quad \text{and} \quad f_{\phi'}(x, \phi, \phi') = \frac{\phi'}{\sqrt{1 + (\phi')^2}}. \tag{7.67}$$

On using (7.67), the Euler equation (7.63) can be integrated immediately to give

$$\phi'(x) = \text{constant} \qquad \text{for} \quad 0 \leq x \leq x_R.$$

The boundary condition at the left hand end is

$$\phi(0) = a,$$

while the boundary conditions (7.66) at the right hand end give

$$\frac{1}{\sqrt{1 + \phi'^2(x_R)}} = \lambda c_1, \qquad \frac{\phi'(x_R)}{\sqrt{1 + \phi'^2(x_R)}} = \lambda c_2.$$

Finally the condition $G(x_R, \phi(x_R)) = 0$ requires that

$$c_1 x_R + c_2 \phi(x_R) + c_3 = 0.$$

Solving the preceding equations leads to the minimizer

$$\phi(x) = (c_2/c_1)x + a \qquad \text{for} \quad 0 \leq x \leq -\frac{c_1(ac_2 + c_3)}{c_1^2 + c_2^2}.$$

7.6 Constrained Minimization

7.6.1 Integral constraints.

Consider a problem of the following general form: find admissible functions $\phi_1(x), \phi_2(x)$ that minimizes

$$F\{\phi_1, \phi_2\} = \int_0^1 f(x,\, \phi_1(x),\, \phi_2(x),\, \phi_1'(x),\, \phi_2'(x))\; dx \tag{7.68}$$

subject to the constraint

$$G(\phi_1, \phi_2) = \int_0^1 f(x,\, \phi_1(x),\, \phi_2(x),\, \phi_1'(x),\, \phi_2'(x))\; dx = 0. \tag{7.69}$$

For reasons of clarity we shall return to the more detailed approach where we introduce parameters $\varepsilon_1, \varepsilon_2$ and functions $\eta_1(x), \eta_1(x)$, rather than following the formal approach using variations $\delta\phi_1(x), \delta\phi_2(x)$. Accordingly, suppose that the pair $\phi_1(x), \phi_2(x)$ is the minimizer. By evaluating F and G on a family of neighboring admissible functions $\phi_1(x)+\varepsilon_1\eta_1(x), \phi_2(x)+\varepsilon_2\eta_2(x)$ we have

$$\begin{aligned} \hat{F}(\varepsilon_1, \varepsilon_2) &= F\{\phi_1(x)+\varepsilon_1\eta_1(x), \phi_2(x)+\varepsilon_2\eta_2(x)\}, \\ \hat{G}(\varepsilon_1, \varepsilon_2) &= G(\phi_1(x)+\varepsilon_1\eta_1(x), \phi_2(x)+\varepsilon_2\eta_2(x)) = 0. \end{aligned} \tag{7.70}$$

If we begin by keeping η_1 and η_2 fixed, this is a classical minimization problem for a function of two variables: we are to minimize the function $\hat{F}(\varepsilon_1, \varepsilon_2)$ with respect to the variables ε_1 and ε_2, subject to the constraint $\hat{G}(\varepsilon_1, \varepsilon_2) = 0$. A necessary condition for this is that $d\hat{F}(\varepsilon_1, \varepsilon_2) = 0$, i.e. that

$$d\hat{F} = \frac{\partial \hat{F}}{\partial \varepsilon_1} d\varepsilon_1 + \frac{\partial \hat{F}}{\partial \varepsilon_2} d\varepsilon_2 = 0, \tag{7.71}$$

for all $d\varepsilon_1, d\varepsilon_2$ that are consistent with the constraint. Because of the constraint, $d\varepsilon_1$ and $d\varepsilon_2$ cannot be varied independently. Instead the constraint requires that they be related by

$$d\hat{G} = \frac{\partial \hat{G}}{\partial \varepsilon_1} d\varepsilon_1 + \frac{\partial \hat{G}}{\partial \varepsilon_2} d\varepsilon_2 = 0. \tag{7.72}$$

If we didn't have the constraint, then (7.71) would imply the usual requirements $\partial\hat{F}/\partial\varepsilon_1 = \partial\hat{F}/\partial\varepsilon_2 = 0$. However when the constraint equation (7.72) holds, (7.71) only requires that

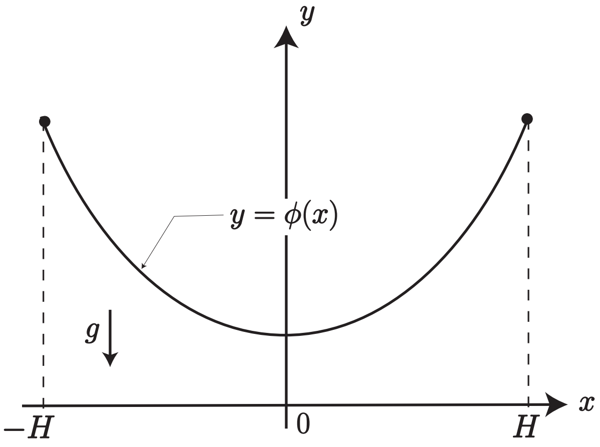
$$\frac{\partial \hat{F}}{\partial \varepsilon_1} = \lambda \frac{\partial \hat{G}}{\partial \varepsilon_1}, \qquad \frac{\partial \hat{F}}{\partial \varepsilon_2} = \lambda \frac{\partial \hat{G}}{\partial \varepsilon_2}, \tag{7.73}$$

for some constant λ, or equivalently

$$\frac{\partial}{\partial \varepsilon_1}(\hat{F} - \lambda \hat{G}) = 0, \qquad \frac{\partial}{\partial \varepsilon_2}(\hat{F} - \lambda \hat{G}) = 0. \tag{7.74}$$

Therefore minimizing $\hat{F}$ subject to the constraint $\hat{G} = 0$ is equivalent to minimizing $\hat{F} - \lambda\hat{G}$ without regard to the constraint; λ is known as a Lagrange multiplier. Proceeding from here on leads to the Euler equation associated with $F - \lambda G$. The presence of the additional unknown parameter λ is balanced by the availability of the constraint equation $G = 0$.

Example: Consider a heavy inextensible cable of mass per unit length m that hangs under gravity. The two ends of the cable are held at the same vertical height, a distance $2H$ apart. The cable has a given length L. We know from physics that the cable adopts a shape that minimizes the potential energy. We are asked to determine this shape.



Let $y = \phi(x)$, $-H \leq x \leq H$, describe an admissible shape of the cable. The potential energy of the cable is determined by integrating $mg\phi$ with respect to the arc length s along the cable which, since $ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + (\phi')^2}\,dx$, is given by

$$V\{\phi\} = \int_0^L mg\phi\,ds = mg\int_{-H}^{H} \phi\sqrt{1 + (\phi')^2}\,dx. \tag{7.75}$$

Since the cable is inextensible, its length

$$\ell\{\phi\} = \int_0^L ds = \int_{-H}^{H} \sqrt{1 + (\phi')^2}\,dx \tag{7.76}$$

must equal L. Therefore we are asked to find a function $\phi(x)$ with $\phi(-H) = \phi(H)$, that minimizes $V\{\phi\}$ subject to the constraint $\ell\{\phi\} = L$. According to the theory developed above, this function must satisfy the Euler equation associated with the functional $V\{\phi\} - \lambda\ell\{\phi\}$ where the Lagrange multiplier λ is a constant. The resulting boundary value problem together with the constraint $\ell = L$ yields the shape of the cable $\phi(x)$.

Calculating the first variation of $V - \lambda mg\ell$, where the constant λ is a Lagrange multiplier, leads to the Euler equation

$$\frac{d}{dx}\left\{(\phi-\lambda)\,\frac{\phi'}{\sqrt{1+(\phi')^2}}\right\} \;-\; \sqrt{1+(\phi')^2} \;=\; 0, \qquad -H < x < H.$$

This can be integrated once to yield

$$\phi' = \sqrt{\frac{(\phi-\lambda)^2}{c^2} \;-\; 1}$$

where c is a constant of integration. Integrating this again leads to

$$\phi(x) = c\cosh[(x+d)/c] + \lambda, \qquad -H < x < H,$$

where d is a second constant of integration. For symmetry, we must have $\phi'(0) = 0$ and therefore $d = 0$. Thus

$$\phi(x) = c\cosh(x/c) + \lambda, \qquad -H < x < H. \tag{7.77}$$

The constant λ in (7.77) is simply a reference height. For example we could take the x-axis to pass through the two pegs in which case $\phi(\pm H) = 0$ and then $\lambda = -c\,\cosh(H/c)$ and so

$$\phi(x) = c\Big[\cosh(x/c) \;-\; \cosh(H/c)\Big], \qquad -H < x < H. \tag{7.78}$$

Substituting (7.78) into the constraint condition $\ell\{\phi\} = L$ with ℓ given by (7.76) yields

$$L = 2c\sinh(H/c). \tag{7.79}$$

Thus in summary, if equation (7.79) can be solved for c, then (7.78) gives the equation describing the shape of the cable.

All that remains is to examine the solvability of (7.79). To this end set $z = H/c$ and $\mu = L/(2H)$. Then we must solve $\mu z = \sinh z$ where $\mu > 1$ is a constant. (The requirement $\mu > 1$ follows from the physical necessity that the distance between the pegs, $2H$, be less than the length of the rope, L.) One can show that as z increases from 0 to ∞, the function $\sinh z - \mu z$ starts from the value 0, decreases monotonically to some finite negative value at some $z = z_* > 0$, and then increases monotonically to ∞. Thus for each $\mu > 0$ the function $\sinh z - \mu z$ vanishes at some unique positive value of z. Consequently (7.79) has a unique root $c > 0$.

7.6.2 Algebraic constraints

Now consider a problem of the following general type: find a pair of admissible functions $\phi_1(x), \phi_2(x)$ that minimizes

$$\int_0^1 f(x, \phi_1, \phi_2, \phi_1', \phi_2')dx$$

subject to the *algebraic constraint*

$$g(x, \phi_1(x), \phi_2(x)) = 0 \qquad \text{for} \quad 0 \le x \le 1.$$

One can show that a necessary condition is that the minimizer should satisfy the Euler equation associated with $f - \lambda g$. In this problem the Lagrange multiplier λ *may be a function of* x.

Example: Consider a conical surface characterized by

$$g(x_1, x_2, x_3) = x_1^2 + x_2^2 - R^2(x_3) = 0, \qquad R(x_3) = x_3 \tan\alpha, \qquad x_3 > 0.$$

Let P $= (p_1, p_2, p_3)$ and Q $= (q_1, q_2, q_3)$, $q_3 > p_3$, be two points on this surface. A smooth wire lies entirely on the conical surface and joins the points P and Q. A bead slides along the wire under gravity, beginning at rest from P. From among all such wires, we are to find the one that gives the minimum travel time.

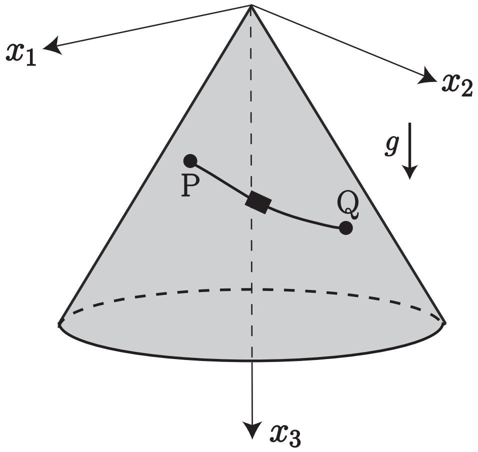


Figure 7.11: A curve that joins the points (p_1, p_2, p_3) to (q_2, q_2, q_3) and lies on the conical surface $x_1^2 + x_2^2 - x_3^2 \tan^2\alpha = 0$.

Suppose that the wire can be described parametrically by $x_1 = \phi_1(x_3), x_2 = \phi_2(x_3)$ for $p_3 \le x_3 \le q_3$. (Not all of the permissible curves can be described this way and so by using this characterization we are limiting ourselves to a subset of all the permited curves.) Since the curve has to lie on the conical surface it is necessary that

$$g(\phi_1(x_3), \phi_2(x_3), x_3) = 0, \qquad p_3 \le x_3 \le q_3. \tag{7.80}$$

The travel time is found by integrating ds/v along the path. The arc length ds along the path is given by

$$ds = \sqrt{dx_1^2 + dx_2^2 + dx_3^2} = \sqrt{(\phi_1')^2 + (\phi_2')^2 + 1}\; dx_3.$$

The conservation of energy tells us that $\frac{1}{2}mv^2(t) - mgx_3(t) = -mgp_3$, or

$$v = \sqrt{2g(x_3 - p_3)}.$$

Therefore the travel time is

$$T\{\phi_1, \phi_2\} \;=\; \int_{p_3}^{q_3} \frac{\sqrt{(\phi_1')^2 + (\phi_2')^2 + 1}}{\sqrt{2g(x_3 - p_3)}}\; dx_3\ .$$

Our task is to minimize $T\{\phi_1, \phi_2\}$ over the set of admissible functions

$$\mathsf{A} = \Big\{ (\phi_1, \phi_2) \;\Big|\; \phi_i : [p_3, q_3] \to \mathbb{R},\ \phi_i \in C^2[p_3, q_3],\ \phi_i(p_3) = p_i,\ \phi_i(q_3) = q_i,\ i = 1, 2 \Big\},$$

subject to the constraint

$$g(\phi_1(x_3), \phi_2(x_3), x_3) = 0, \qquad p_3 \le x_3 \le q_3.$$

According to the theory developed the solution is given by solving the Euler equations associated with $f - \lambda(x_3) g$ where

$$f(x_3, \phi_1, \phi_2, \phi_1', \phi_2') = \frac{\sqrt{(\phi_1')^2 + (\phi_2')^2 + 1}}{\sqrt{2g(x_3 - p_3)}} \quad \text{and} \quad g(x_1, x_2, x_3) = x_1^2 + x_2^2 - x_3^2 \tan^2 \alpha,$$

subject to the prescribed conditions at the ends and the constraint $g(x_1, x_2, x_3) = 0$.

7.6.3 Differential constraints

Now consider a problem of the following general type: find a pair of admissible functions $\phi_1(x), \phi_2(x)$ that minimizes

$$\int_0^1 f(x, \phi_1, \phi_2, \phi_1', \phi_2') dx$$

subject to the *differential equation constraint*

$$g(x, \phi_1(x), \phi_2(x), \phi_1'(x), \phi_2'(x)) = 0, \qquad \text{for} \quad 0 \le x \le 1.$$

Suppose that the constraint is not integrable, i.e. suppose that there does not exist a function $h(x, \phi_1(x), \phi_2(x))$ such that $g = dh/dx$. (In dynamics, such constraints are called non-holonomic.) One can show that it is necessary that the minimizer satisfy the Euler equation associated with $f - \lambda g$. In these problems, the Lagrange multiplier λ *may be a function of* x.

Example: Determine functions $\phi_1(x)$ and $\phi_2(x)$ that minimize

$$\int_0^1 f(x, \phi_1, \phi_1', \phi_2')dx$$

over an admissible set of functions subject to the non-holonomic constraint

$$g(x, \phi_1, \phi_2, \phi_1', \phi_2') = \phi_2 - \phi_1' = 0, \qquad \text{for} \quad 0 \le x \le 1. \tag{7.81}$$

According to the theory above, the minimizers satisfy the Euler equations

$$\frac{d}{dx}\left[\frac{\partial h}{\partial \phi_1'}\right] - \frac{\partial h}{\partial \phi_1} = 0, \qquad \frac{d}{dx}\left[\frac{\partial h}{\partial \phi_2'}\right] - \frac{\partial h}{\partial \phi_2} = 0 \qquad \text{for} \quad 0 < x < 1, \tag{7.82}$$

where $h = f - \lambda g$. On substituting for f and g, these Euler equations reduce to

$$\frac{d}{dx}\left[\frac{\partial f}{\partial \phi_1'} + \lambda\right] - \frac{\partial f}{\partial \phi_1} = 0, \qquad \frac{d}{dx}\left[\frac{\partial f}{\partial \phi_2'}\right] + \lambda = 0 \qquad \text{for} \quad 0 < x < 1. \tag{7.83}$$

Thus the functions $\phi_1(x), \phi_2(x), \lambda(x)$ are determined from the three differential equations (7.81), (7.83).

Remark: Note by substituting the constraint into the integrand of the functional that we can equivalently pose this problem as one for determining the function $\phi_1(x)$ that minimizes

$$\int_0^1 f(x, \phi_1, \phi_1', \phi_1'')dx$$

over an admissible set of functions.

7.7 Piecewise smooth minimizers. Weirstrass-Erdman corner conditions.

In order to motivate the discussion to follow, first consider the problem of minimizing the functional

$$F\{\phi\} = \int_0^1 ((\phi')^2 - 1)^2 dx \tag{7.84}$$

over functions ϕ with $\phi(0) = \phi(1) = 0$.

This is apparently a problem of the classical type where in the present case we are to minimize the integral of $f(x, \phi, \phi') = \big[(\phi')^2 - 1\big]^2$ with respect to x over the interval $[0, 1]$. Assuming that the class of admissible functions are those that are $C^1[0, 1]$ and satisfy $\phi(0) = \phi(1) = 0$, the minimizer must necessarily satisfy the Euler equation $\frac{d}{dx}(\partial f/\partial\phi') - (\partial f/\partial\phi) = 0$. In the present case this specializes to $2[\,(\phi')^2 - 1](2\phi') =$ constant for $0 \leq x \leq 1$, which in turn gives $\phi'(x) =$ constant for $0 \leq x \leq 1$. On enforcing the boundary conditions $\phi(0) = \phi(1) = 0$, this gives $\phi(x) = 0$ for $0 \leq x \leq 1$. This is an extremizer of $F\{\phi\}$ over the class of admissible functions under consideration. It is readily seen from (7.84) that the value of F at this particular function $\phi(x) = 0$ is $F = 1$

Note from (7.84) that $F \geq 0$. It is natural to wonder whether there is a function $\phi_*(x)$ that gives $F\{\phi_*\} = 0$. If so, ϕ_* would be a minimizer. If there is such a function ϕ_*, we know that it cannot belong to the class of admissible functions considered above, since if it did, we would have found it from the preceding calculation. Therefore if there is a function ϕ_* of this type, it does not belong to the set of functions A. The functions in A were required to be $C^1[0, 1]$ and to vanish at the two ends $x = 0$ and $x = 1$. Since $\phi_* \notin \mathsf{A}$ it must not satisfy one or both of these two conditions. The problem statement requires that the boundary conditions must hold. Therefore it must be true that ϕ is not as smooth as $C^1[0, 1]$.

If there is a function ϕ_* such that $F\{\phi_*\} = 0$, it follows from the nonnegative character of the integrand in (7.84) that the integrand itself should vanish almost everywhere in $[0, 1]$. This requires that $\phi'(x) = \pm 1$ almost everywhere in $[0, 1]$. The piecewise linear function

$$\phi_*(x) = \begin{cases} x & \text{for} \quad 0 \leq x \leq 1/2, \\ \\ (1 - x) & \text{for} \quad 1/2 \leq x \leq 1, \end{cases} \tag{7.85}$$

has this property. It is continuous, is piecewise C^1, and gives $F\{\phi_*\} = 0$. Moreover $\phi_*(x)$ satisfies the Euler equation except at $x = 1/2$.

But is it legitimate for us to consider piecewise smooth functions? If so are there are any restrictions that we must enforce? Physical problems involving discontinuities in certain physical fields or their derivatives often arise when, for example, the problem concerns an interface separating two different mateirals. A specific example will be considered below.

7.7.1 Piecewise smooth minimizer with non-smoothness occuring at a prescribed location.

Suppose that we wish to extremize the functional

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi')dx$$

over some suitable set of admissible functions, and suppose further that we know that the extremal $\phi(x)$ is continuous but has a discontinuity in its slope at $x = s$: i.e. $\phi'(s-) \neq \phi'(s+)$ where $\phi'(s\pm)$ denotes the limiting values of $\phi'(s \pm \varepsilon)$ as $\varepsilon \to 0$. Thus the set of admissible functions is composed of all functions that are smooth on either side of $x = s$, that are continuous at $x = s$, and that satisfy the given boundary conditions $\phi(0) = \phi_0, \phi(1) = \phi_1$:

$$\mathsf{A} = \{\phi(\cdot)\big|\phi : [0,1] \to \mathbb{R},\ \phi \in C^1([0,s) \cup (s,1]),\ \phi \in C[0,1],\ \phi(0) = \phi_0,\ \phi(1) = \phi_1\}\ .$$

Observe that an admissible function is required to be continuous on $[0,1]$, required to have a continuous first derivative on either side of $x = s$, and its first derivative is permitted to have a jump discontinuity at a *given* location $x = s$.

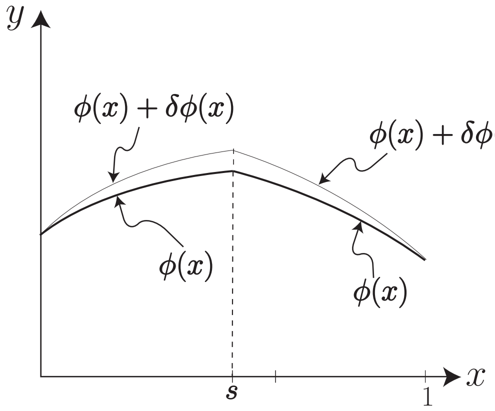


Figure 7.12: Extremal $\phi(x)$ and a neighboring test function $\phi(x) + \delta\phi(x)$ both with kinks at $x = s$.

Suppose that F is extremized by a function $\phi(x) \in \mathsf{A}$ and suppose that this extremal has a jump discontinuity in its first derivative at $x = s$. Let $\delta\phi(x)$ be an admissible variation which means that the neighboring function $\phi(x) + \delta(x)$ is also in A which means that it is C^1 on $[0, s) \cup (s, 1]$ and (may) have a jump discontinuity in its first derivative at the location $x = s$; see Figure 7.12. This implies that

$$\delta\phi(x) \in C([0,1]), \qquad \delta\phi(x) \in C^1([0,s) \cup (s,1]), \qquad \delta\phi(0) = \delta\phi(1) = 0.$$

In view of the lack of smoothness at $x = s$ it is convenient to split the integral into two parts and write

$$F\{\phi\} = \int_0^s f(x, \phi, \phi')dx + \int_s^1 f(x, \phi, \phi')dx,$$

and

$$F\{\phi + \delta\phi\} = \int_0^s f(x, \phi + \delta\phi, \phi' + \delta\phi')dx + \int_s^1 f(x, \phi + \delta\phi, \phi' + \delta\phi')dx.$$

Upon calculating δF, which by definition equals $F\{\phi + \delta\phi\} - F\{\phi\}$ upto terms linear in $\delta\phi$, and setting $\delta F = 0$, we obtain

$$\int_0^s \left(f_\phi \delta\phi + f_{\phi'}\delta\phi'\right) dx + \int_s^1 \left(f_\phi \delta\phi + f_{\phi'}\delta\phi'\right) dx = 0.$$

Integrating the terms involving $\delta\phi'$ by parts leads to

$$\int_0^s \left[f_\phi - \frac{d}{dx}\left(f_{\phi'}\right)\right] \delta\phi \; dx \; + \int_s^1 \left[f_\phi - \frac{d}{dx}\left(f_{\phi'}\right)\right] \delta\phi \; dx \; + \left[\frac{\partial f}{\partial \phi'}\delta\phi\right]_{x=0}^{s-} + \left[\frac{\partial f}{\partial \phi'}\delta\phi\right]_{x=s+}^{1} = 0.$$

However, since $\delta\phi(0) = \delta\phi(1) = 0$, this simplifies to

$$\int_0^1 \left[\frac{\partial f}{\partial \phi} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\right] \delta\phi(x) \; dx \; + \left(\left.\frac{\partial f}{\partial \phi'}\right|_{x=s-} - \left.\frac{\partial f}{\partial \phi'}\right|_{x=s+}\right) \delta\phi(s) = 0.$$

First, if we limit attention to variations that are such that $\delta\phi(s) = 0$, the second term in the equation above vanishes, and only the integral remains. Since $\delta\phi(x)$ can be chosen arbitrarily for all $x \in (0,1)$, $x \neq s$, this implies that the term within the brackets in the integrand must vanish at each of these x's. This leads to the Euler equation

$$\frac{\partial f}{\partial \phi} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right) = 0 \qquad \text{for} \quad 0 < x < 1, \; x \neq s.$$

Second, when this is substituted back into the equation above it, the integral now disappears. Since the resulting equation must hold for all variations $\delta\phi(s)$, it follows that we must have

$$\left.\frac{\partial f}{\partial \phi'}\right|_{x=s-} = \left.\frac{\partial f}{\partial \phi'}\right|_{x=s+}$$

at $x = s$. This is a "matching condition" or "jump condition" that relates the solution on the left of $x = s$ to the solution on its right. The matching condition shows that even though ϕ' has a jump discontinuity at $x = s$, the quantity $\partial f/\partial\phi'$ is continuous at this point.

Thus in summary an extremal ϕ must obey the following boundary value problem:

$$\left.\begin{aligned}
\frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right) - \frac{\partial f}{\partial \phi} = 0 \qquad \text{for} \quad 0 < x < 1, \ x \neq s,\\
\phi(0) = \phi_0,\\
\phi(1) = \phi_1,\\
\left.\frac{\partial f}{\partial \phi'}\right|_{x=s-} = \left.\frac{\partial f}{\partial \phi'}\right|_{x=s+} \quad \text{at } x = s.
\end{aligned}\right\} \tag{7.86}$$

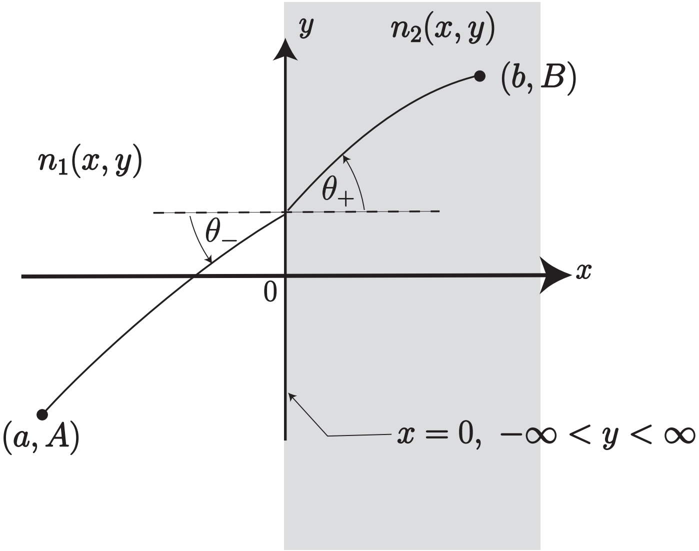


Figure 7.13: Ray of light in a two-phase material.

Example: Consider a two-phase material that occupies all of x, y-space. The material occupying $x < 0$ is different to the material occupying $x > 0$ and so $x = 0$ is the interface between the two materials. In particular, suppose that the refractive indices of the materials occupying $x < 0$ and $x > 0$ are $n_1(x, y)$ and $n_2(x, y)$ respectively; see Figure 7.13. We are asked to determine the path $y = \phi(x), a \leq x \leq b$, followed by a ray of light travelling from a point (a, A) in the left half-plane to the point (b, B) in the right half-plane. In particular, we are to determine the conditions at the point where the ray crosses the interface between the two media.

According to Fermat's principle, a ray of light travelling between two given points follows the path that it can traverse in the shortest possible time. Also, we know that light travels at a speed $c/n(x,y)$ where $n(x,y)$ is the index of refraction at the point (x,y). Thus the transit time is determined by integrating n/c along the path followed by the light, which, since $ds = \sqrt{1+(\phi')^2}\ dx$ can be written as

$$T\{\phi\} = \int_a^b \frac{1}{c}\ n(x,\phi(x))\sqrt{1+(\phi')^2}\ dx.$$

Thus the problem at hand is to determine ϕ that minimizes the functional $T\{\phi\}$ over the set of admissible functions

$$\mathsf{A} = \{\phi(\cdot)\big|\phi \in C[a,b], \phi \in C^1([a,0)\cup(0,b]), \phi(a) = A, \phi(b) = B\}.$$

Note that this set of admissible functions allows the path followed by the light to have a kink at $x=0$ even though the path is continuous.

The functional we are asked to minimize can be written in the standard form

$$T\{\phi\} = \int_a^b f(x,\phi,\phi')\ dx \qquad \text{where} \qquad f(x,\phi,\phi') = \frac{n(x,\phi)}{c}\ \sqrt{1+(\phi')^2}\ .$$

Therefore

$$\frac{\partial f}{\partial \phi'} = \frac{n(x,\phi)}{c}\ \frac{\phi'}{\sqrt{1+(\phi')^2}}$$

and so the matching condition at the kink at $x=0$ requires that

$$\frac{n}{c}\ \frac{\phi'}{\sqrt{1+(\phi')^2}} \quad \text{be continuous at} \quad x = 0.$$

Observe that, if θ is the angle made by the ray of light with the x-axis at some point along its path, then $\tan\theta = \phi'$ and so $\sin\theta = \phi'/\sqrt{1+(\phi')^2}$. Therefore the matching condition requires that $n\sin\theta$ be continuous, or

$$n_+ \sin\theta_+ = n_- \sin\theta_-$$

where $n_\pm$ and $\theta_\pm$ are the limiting values of $n(x,\phi(x))$ and $\theta(x)$ as $x \to 0\pm$. This is Snell's well-known law of refraction.

7.7.2 Piecewise smooth minimizer with non-smoothness occuring at an unknown location

Suppose again that we wish to extremize the functional

$$F(\phi) = \int_0^1 f(x, \phi, \phi')\, dx$$

over the admissible set of functions

$$\mathsf{A} = \{\phi(\cdot) : \phi : [0,1] \to \mathbb{R}, \phi \in C[0,1], \phi \in C^1_p[0,1], \phi(0) = a, \phi(1) = b\}$$

Just as before, the admissible functions are continuous and have a piecewise continuous first derivative. However in contrast to the preceding case, if there is discontinuity in the first derivative of ϕ at some location $x = s$, the location s is *not known a priori* and so is also to be determined.

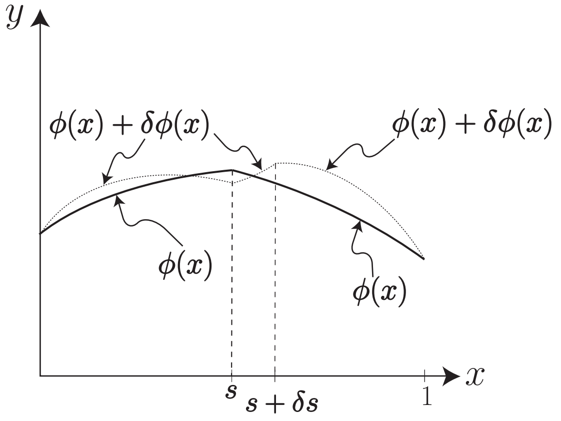


Figure 7.14: Extremal $\phi(x)$ with a kink at $x = s$ and a neighboring test function $\phi(x) + \delta\phi(x)$ with kinks at $x = s$ and $s + \delta s$.

Suppose that F is extremized by the function $\phi(x)$ and that it has a jump discontinuity in its first derivative at $x = s$; (we shall say that ϕ has a "kink" at $x = s$). Suppose further that ϕ is C^1 on either side of $x = s$. Consider a variation $\delta\phi(x)$ that vanishes at the two ends $x = 0$ and $x = 1$, is continuous on $[0,1]$, is C^1 on $[0,1]$ except at $x = s + \delta s$ where it has a jump discontinuity in its first derivative:

$$\delta\phi \in C[0,1] \cup C^1[0, s + \delta s) \cup C^1(s + \delta s, 1], \quad \delta\phi(0) = \delta\phi(1) = 0.$$

Note that $\phi(x) + \delta\phi(x)$ has kinks at both $x = s$ and $x = s + \delta s$. Note further that we have varied the function $\phi(x)$ and the location of the kink s. See Figure 7.14.

Since the extremal $\phi(x)$ has a kink at $x = s$ it is convenient to split the integral and express $F\{\phi\}$ as

$$F\{\phi\} \;=\; \int_0^s f(x, \phi, \phi')dx + \int_s^1 f(x, \phi, \phi')dx.$$

Similarly since the the neigboring function $\phi(x) + \delta(x)$ has kinks at $x = s$ and $x = s + \delta s$, it is convenient to express $F\{\phi + \delta\phi\}$ by splitting the integral into three terms as follows:

$$\begin{aligned} F\{\phi + \delta\phi\} \;&=\; \int_0^s f(x, \phi + \delta\phi, \phi' + \delta\phi')dx + \int_s^{s+\delta s} f(x, \phi + \delta\phi, \phi' + \delta\phi')dx \\ &+\; \int_{s+\delta s}^1 f(x, \phi + \delta\phi, \phi' + \delta\phi')dx. \end{aligned}$$

We can now calculate the first variation δF which, by definition, equals $F\{\phi + \delta\phi\} - F\{\phi\}$ upto terms linear in $\delta\phi$. Calculating δF in this way and setting the result equal to zero, leads after integrating by parts, to

$$\int_0^1 A\,\delta\phi(x)dx \;+\; B\,\delta\phi(s) \;+\; C\,\delta s = 0,$$

where

$$\begin{aligned} A \;&=\; \frac{\partial f}{\partial \phi} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right), \\ B \;&=\; \left(\frac{\partial f}{\partial \phi'}\right)_{x=s-} - \left(\frac{\partial f}{\partial \phi'}\right)_{x=s+}, \qquad (7.87) \\ C \;&=\; \left(f - \phi'\frac{\partial f}{\partial \phi'}\right)_{x=s-} - \left(f - \phi'\frac{\partial f}{\partial \phi'}\right)_{x=s+}. \end{aligned}$$

By the arbitrariness of the variations above, it follows in the usual way that A, B and C all must vanish. This leads to the usual Euler equation on $(0, s) \cup (s, 1)$, and the following two additional requirements at $x = s$:

$$\left.\frac{\partial f}{\partial \phi'}\right|_{s-} \;=\; \left.\frac{\partial f}{\partial \phi'}\right|_{s+}, \qquad (7.88)$$

$$\left.\left(f - \phi'\frac{\partial f}{\partial \phi'}\right)\right|_{s-} = \left.\left(f - \phi'\frac{\partial f}{\partial \phi'}\right)\right|_{s+}. \qquad (7.89)$$

The two matching conditions (or jump conditions) (7.88) and (7.89) are known as the *Wierstrass-Erdmann corner conditions* (the term “corner” referring to the “kink” in ϕ). Equation (7.88) is the same condition that was derived in the preceding subsection.

Example: Find the extremals of the functional

$$F(\phi) = \int_0^4 f(x,\phi,\phi')\ dx = \int_0^4 (\phi'-1)^2(\phi'+1)^2 dx$$

over the set of piecewise smooth functions subject to the end conditions $\phi(0) = 0, \phi(4) = 2$. For simplicity, restrict attention to functions that have at most one point at which ϕ' has a discontinuity.

Here

$$f(x,\phi,\phi') = \big[(\phi')^2 - 1\big]^2 \tag{7.90}$$

and therefore on differentiating f,

$$\frac{\partial f}{\partial \phi'} = 4\phi'\big[(\phi')^2-1\big], \qquad \frac{\partial f}{\partial \phi} = 0. \tag{7.91}$$

Consequently the Euler equation (at points of smoothness) is

$$\frac{d}{dx}f_{\phi'} - f_\phi \ = \ \frac{d}{dx}\ \big[4\phi'\big((\phi')^2-1\big)\big] = 0. \tag{7.92}$$

First, consider an extremal that is smooth everywhere. (Such an extremal might not, of course, exist.) In this case the Euler equation (7.92) holds on the entire interval $(0,4)$ and so we conclude that $\phi'(x) =$ constant for $0 \le x \le 4$. On integrating this and using the boundary conditions $\phi(0) = 0, \phi(4) = 2$, we find that $\phi(x) = x/2,\ 0 \le x \le 4$, is a smooth extremal. In order to compare this with what follows, it is helpful to call this, say, ϕ_0. Thus

$$\phi_o(x) = \frac{1}{2}x \qquad \text{for} \quad 0 \le x \le 4,$$

is a smooth extremal of F.

Next consider a piecewise smooth extremizer of F which has a kink at some location $x = s$; the value of $s \in (0,4)$ is not known a priori and is to be determined. (Again, such an extremal might not, of course, exist.) The Euler equation (7.92) now holds *on either side* of $x = s$ and so we find from (7.92) that $\phi' = c =$ constant on $(0,s)$ and $\phi' = d =$ constant on $(s,4)$ where $c \ne d$; (if $c = d$ there would be no kink at $x = s$ and we have already dealt with this case above). Thus

$$\phi'(x) = \begin{cases} c & \text{for} \quad 0 < x < s, \\ \\ d & \text{for} \quad s < x < 4. \end{cases}$$

Integrating this, separately on $(0, s)$ and $(s, 4)$, and enforcing the boundary conditions $\phi(0) = 0$, $\phi(4) = 2$, leads to

$$\phi(x) = \begin{cases} cx & \text{for} \quad 0 \le x \le s, \\ d(x-4)+2 & \text{for} \quad s \le x \le 4. \end{cases} \tag{7.93}$$

Since ϕ is required to be continuous, we must have $\phi(s-) = \phi(s+)$ which requires that $cs = d(s-4)+2$ whence

$$s = \frac{2-4d}{c-d} \ . \tag{7.94}$$

Note that s would not exist if $c = d$.

All that remains is to find c and d, and the two Weirstrass-Erdmann corner conditions (7.88), (7.89) provide us with the two equations for doing this. From (7.90), (7.91) and (7.93),

$$\frac{\partial f}{\partial \phi'} = \begin{cases} 4c(c^2-1) & \text{for} \quad 0 < x < s, \\ 4d(d^2-1) & \text{for} \quad s < x < 4. \end{cases}$$

and

$$f - \phi' \frac{\partial f}{\partial \phi'} = \begin{cases} -(c^2-1)(1+3c^2) & \text{for} \quad 0 < x < s, \\ -(d^2-1)(1+3d^2) & \text{for} \quad s < x < 4. \end{cases}$$

Therefore the Weirstrass-Erdmann corner conditions (7.88) and (7.89), which require respectively the continuity of $\partial f/\partial \phi'$ and $f - \phi' \partial f/\partial \phi'$ at $x = s$, give us the pair of simultaneous equations

$$\left. \begin{array}{rcl} c(c^2-1) & = & d(d^2-1), \\ (c^2-1)(1+3c^2) & = & (d^2-1)(1+3d^2). \end{array} \right\}$$

Keeping in mind that $c \neq d$ and solving these equations leads to the two solutions:

$$c = 1, \ d = -1, \qquad \text{and} \qquad c = -1, \ d = 1.$$

Corresponding to the former we find from (7.94) that $s = 3$, while the latter leads to $s = 1$. Thus from (7.93) there are two piecewise smooth extremals $\phi_1(x)$ and $\phi_2(x)$ of the assumed form:

$$\phi_1(x) = \begin{cases} x & \text{for} \quad 0 \le x \le 3, \\ -x+6 & \text{for} \quad 3 \le x \le 4. \end{cases}$$

$$\phi_2(x) = \begin{cases} -x & \text{for} \quad 0 \le x \le 1, \\ x-2 & \text{for} \quad 1 \le x \le 4. \end{cases}$$

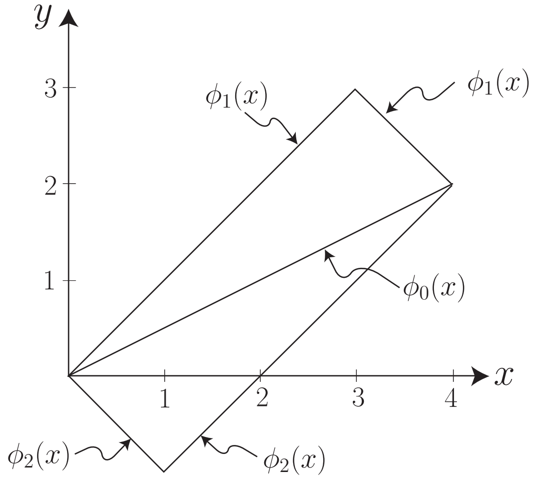


Figure 7.15: Smooth extremal $\phi_0(x)$ and piecewise smooth extremals $\phi_1(x)$ and $\phi_2(x)$.

Figure 7.15 shows graphs of ϕ_o, ϕ_1 and ϕ_2. By evaluating the functional F at each of the extremals ϕ_0, ϕ_1 and ϕ_2, we find

$$F\{\phi_o\} = 9/4, \qquad F\{\phi_1\} = F\{\phi_2\} = 0.$$

Remark: By inspection of the given functional

$$F(\phi) = \int_0^4 \left[(\phi')^2 - 1\right]^2 dx,$$

it is clear that (a) $F \geq 0$, and (b) $F = 0$ if and only if $\phi' = \pm 1$ everywhere (except at isolated points where ϕ' may be undefined). The extremals ϕ_1 and ϕ_2 have this property and therefore correspond to absolute minimizers of F.

7.8 Generalization to higher dimensional space.

In order to help motivate the way in which we will approach higher-dimensional problems (which will in fact be entirely parallel to the approach we took for one-dimensional problems) we begin with some preliminary observations.

First, consider the *one-dimensional* variational problem of minimizing a functional

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi', \phi'')\ dx$$

on a set of suitably smooth functions with no prescribed boundary conditions at either end. The analogous two-dimensional problem would be to consider a set of suitably smooth functions $\phi(x, y)$ defined on a domain $\mathcal{D}$ of the x, y-plane and to minimize a given functional

$$F\{\phi\} = \int_{\mathcal{D}} f(x,\, y,\, \phi,\, \partial\phi/\partial x,\, \partial\phi/\partial y,\, \partial^2\phi/\partial x^2,\, \partial^2\phi/\partial x\partial y,\, \partial^2\phi/\partial y^2)\; dA$$

over this set of functions with no boundary conditions prescribed anywhere on the boundary $\partial\mathcal{D}$.

In deriving the Euler equation in the one-dimensional case our strategy was to exploit the fact that the variation $\delta\phi(x)$ was arbitrary in the interior $0 < x < 1$ of the domain. This motivated us to express the integrand in the form of some quantity A (independent of any variations) multiplied by $\delta\phi(x)$. Then, the arbitrariness of $\delta\phi$ allowed us to conclude that A must vanish on the entire domain. We approach two-dimensional problems similarly and our strategy will be to exploit the fact that $\delta\phi(x, y)$ is arbitrary in the interior of $\mathcal{D}$ and so we attempt to express the integrand as some quantity A that is independent of any variations multiplied by $\delta\phi$. Similarly concerning the boundary terms, in the one-dimensional case we were able to exploit the fact that $\delta\phi$ and its derivative $\delta\phi'$ are arbitrary at the boundary points $x = 0$ and $x = 1$, and this motivated us to express the boundary terms as some quantity B that is independent of any variations multiplied by $\delta\phi(0)$, another quantity C independent of any variations multiplied by $\delta\phi'(0)$, and so on. We approach two-dimensional problems similarly and our strategy for the boundary terms is to exploit the fact that $\delta\phi$ and its normal derivative $\partial(\delta\phi)/\partial n$ are arbitrary on the boundary $\partial\mathcal{D}$. Thus the goal in our calculations will be to express the boundary terms as some quantity independent of any variations multiplied by $\delta\phi$, another quantity independent of any variations multiplied by $\partial(\delta\phi)/\partial n$ etc. Thus in the two-dimensional case our strategy will be to take the first variation of F and carry out appropriate calculations that lead us to an equation of the form

$$\delta F = \int_{\mathcal{D}} A\, \delta\phi(x, y)\, dA \;+\; \int_{\partial\mathcal{D}} B\, \delta\phi(x, y)\, ds + \int_{\partial\mathcal{D}} C\, \Big(\frac{\partial}{\partial n}(\delta\phi(x, y))\Big)\, ds \;=\; 0 \qquad (7.95)$$

where A, B, C are independent of $\delta\phi$ and its derivatives and the latter two integrals are on the boundary of the domain $\mathcal{D}$. We then exploit the arbitrariness of $\delta\phi(x, y)$ on the interior of the domain of integration, and the arbitrariness of $\delta\phi$ and $\partial(\delta\phi)/\partial n$ on the boundary ∂D to conclude that the minimizer must satisfy the partial differential equation $A = 0$ for $(x, y) \in \mathcal{D}$ and the boundary conditions $B = C = 0$ on ∂D.

Next, recall that one of the steps involved in calculating the minimizer of a one-dimensional problem is integration by parts. This converts a term that is an integral over $[0, 1]$ into terms that are only evaluated on the boundary points $x = 0$ and 1. The analog of this in higher dimensions is carried out using the divergence theorem, which in two-dimensions reads

$$\int_{\mathcal{D}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right) dA = \int_{\partial\mathcal{D}} (P n_x + Q n_y)\, ds \tag{7.96}$$

which expresses the left hand side, which is an integral over $\mathcal{D}$, in a form that only involves terms on the boundary. Here n_x, n_y are the components of the unit normal vector $\mathbf{n}$ on $\partial\mathcal{D}$ that points out of $\mathcal{D}$. Note that in the special case where $P = \partial\chi/\partial x$ and $Q = \partial\chi/\partial y$ for some $\chi(x, y)$ the integrand of the right hand side is $\partial\chi/\partial n$.

Remark: The derivative of a function $\phi(x, y)$ in a direction corresponding to a unit vector $\mathbf{m}$ is written as $\partial\phi/\partial m$ and defined by $\partial\phi/\partial m = \boldsymbol{\nabla}\phi \cdot \mathbf{m} = (\partial\phi/\partial x)\, m_x + \partial\phi/\partial y)\, m_y$ where m_x and m_y are the components of $\mathbf{m}$ in the x- and y-directions respectively. On the boundary $\partial\mathcal{D}$ of a two dimensional domain $\mathcal{D}$ we frequently need to calculate the derivative of ϕ in directions $\mathbf{n}$ and $\mathbf{s}$ that are normal and tangential to $\partial\mathcal{D}$. In vector form we have

$$\boldsymbol{\nabla}\phi = \frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} = \frac{\partial\phi}{\partial n}\mathbf{n} + \frac{\partial\phi}{\partial s}\mathbf{s}$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors in the x- and y-directions. Recall also that a function $\phi(x, y)$ and its tangential derivative $\partial\phi/\partial s$ along the boundary $\partial\mathcal{D}$ are *not* independent of each other in the following sense: if one knows the values of ϕ along ∂D one can differentiate ϕ along the boundary to get $\partial\phi/\partial s$; and conversely if one knows the values of $\partial\phi/\partial s$ along ∂D one can integrate it along the boundary to find ϕ to within a constant. This is why equation (7.95) does not involve a term of the form $E\,\partial(\delta\phi)/\partial s$ integrated along the boundary ∂D since it can be rewritten as the integral of $-(\partial E/\partial s)\,\delta\phi$ along the boundary

Example 1: A stretched membrane. A stretched flexible membrane occupies a regular region $\mathcal{D}$ of the x, y-plane. A pressure $p(x, y)$ is applied normal to the surface of the membrane in the negative z-direction. Let $u(x, y)$ be the resulting deflection of the membrane in the z-direction. The membrane is fixed along its entire edge $\partial\mathcal{D}$ and so

$$u = 0 \qquad \text{for} \quad (x, y) \in \partial\mathcal{D}. \tag{7.97}$$

One can show that the potential energy Φ associated with any deflection u that is compatible with the given boundary condition is

$$\Phi\{u\} = \int_{\mathcal{D}} \frac{1}{2} \left|\boldsymbol{\nabla} u\right|^2 dA - \int_{\mathcal{D}} p u\, dA$$

where we have taken the relevant stiffness of the membrane to be unity. The actual deflection of the membrane is the function that minimizes the potential energy over the set of test functions

$$\mathsf{A} = \{u \;\big|\; u \in C^2(\mathcal{D}), u = 0 \text{ for } (x,y) \in \partial\mathcal{D}\}.$$

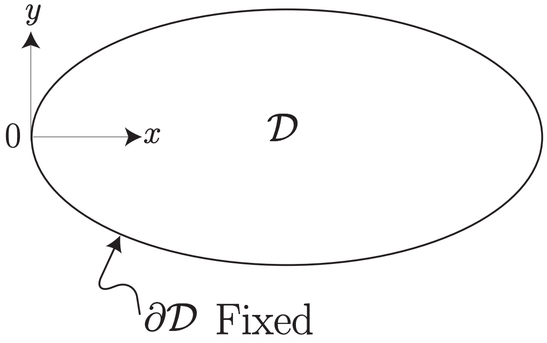


Figure 7.16: A stretched elastic membrane whose mid-plane occupies a region $\mathcal{D}$ of the x, y-plane and whose boundary $\partial\mathcal{D}$ is fixed. The membrane surface is subjected to a pressure loading $p(x, y)$ that acts in the negative z-direction.

Since

$$\Phi\{u\} = \int_{\mathcal{D}} \frac{1}{2}(u_{,x}u_{,x} + u_{,y}u_{,y})\,dA - \int_{\mathcal{D}} pu\,dA,$$

its first variation is

$$\delta\Phi = \int_{\mathcal{D}} (u_{,x}\delta u_{,x} + u_{,y}\delta u_{,y})\,dA - \int_{\mathcal{D}} p\delta u\,dA,$$

where an admissible variation $\delta u(x, y)$ vanishes on $\partial\mathcal{D}$. Here we are using the notation that a comma followed by a subscript denotes partial differentiation with respect to the corresponding coordinate, for example $u_{,x} = \partial u/\partial x$ and $u_{,xy} = \partial^2 u/\partial x\partial y$. In order to make use of the divergence theorem and convert the area integral into a boundary integral we must write the integrand so that it involves terms of the form $(\ldots)_{,x} + (\ldots)_{,y}$; see (7.96). This suggests that we rewrite the preceding equation as

$$\delta\Phi = \int_{\mathcal{D}} \Big((u_{,x}\delta u)_{,x} + (u_{,y}\delta u)_{,y} - (u_{,xx} + u_{,yy})\delta u\Big)\,dA - \int_{\mathcal{D}} p\delta u\,dA,$$

or equivalently as

$$\delta\Phi = \int_{\mathcal{D}} \Big((u_{,x}\delta u)_{,x} + (u_{,y}\delta u)_{,y}\Big)\,dA - \int_{\mathcal{D}} \Big(u_{,xx} + u_{,yy} + p\Big)\delta u\,dA.$$

By using the divergence theorem on the first integral we get

$$\delta\Phi = \int_{\partial\mathcal{D}} \Big(u_{,x}n_x + u_{,y}n_y\Big)\delta u\, ds \;-\; \int_{\mathcal{D}} (u_{,xx} + u_{,yy} + p)\,\delta u\, dA$$

where $\mathbf{n}$ is a unit outward normal along $\partial\mathcal{D}$. We can write this equivalently as

$$\delta\Phi = \int_{\partial\mathcal{D}} \frac{\partial u}{\partial n}\,\delta u\, ds \;-\; \int_{\mathcal{D}} \big(\nabla^2 u + p\big)\,\delta u\, dA. \tag{7.98}$$

Since the variation δu vanishes on $\partial\mathcal{D}$ the first integral drops out and we are left with

$$\delta\Phi = \;-\; \int_{\mathcal{D}} \big(\nabla^2 u + p\big)\;\delta u\, dA \tag{7.99}$$

which must vanish for all admissible variations $\delta u(x, y)$. Thus the minimizer satisfies the partial differential equation

$$\nabla^2 u + p = 0 \qquad \text{for} \quad (x, y) \in \mathcal{D}$$

which is the Euler equation in this case that is to be solved subject to the prescribed boundary condition (7.97). Note that if some part of the boundary of $\mathcal{D}$ had not been fixed, then we would not have $\delta u = 0$ on that part in which case (7.98) and (7.99) would yield the natural boundary condition $\partial\phi/\partial n = 0$ on that segment.

NNN Show the calculations for just one term w_{xx}^2. Include $nu \neq 0$ in text.NNN

NNN Check signs of terms and sign conventionNNN

Example 2: The Kirchhoff theory of plates. We consider the bending of a thin plate according to the so-called Kirchhoff theory. Solely for purposes of mathematical simplicity we shall assume that the Poisson ratio ν of the material is zero. A discussion of the case $\nu \neq 0$ can be found in many books, for example, in "Energy & Finite Elements Methods in Structural Mechanic" by I.H. Shames & C.L. Dym. When $\nu = 0$ the plate bending stiffness $D = Et^3/12$ where E is the Young's modulus of the material and t is the thickness of the plate. The mid-plane of the plate occupies a domain of the x, y-plane and $w(x, y)$ denotes the deflection (displacement) of a point on the mid-plane in the z-direction. The basic constitutive relationships of elastic plate theory relate the internal moments M_x, M_y, M_{xy}, M_{yx} (see Figure 7.17) to the second derivatives of the displacement field $w_{,xx}, w_{,yy}, w_{,xy}$ by

$$M_x = -Dw_{,xx}, \quad M_y = -Dw_{,yy}, \quad M_{xy} = M_{yx} = -Dw_{,xy}, \tag{7.100}$$

where a comma followed by a subscript denotes partial differentiation with respect to the corresponding coordinate and D is the plate bending stiffness; and the shear forces in the plate are given by

$$V_x = -D(\nabla^2 w)_{,x}, \qquad V_y = -D(\nabla^2 w)_{,y}. \tag{7.101}$$

The elastic energy per unit volume of the plate is given by

$$\frac{1}{2}\big(M_x w_{,xx} + M_y w_{,yy} + M_{xy} w_{,xy} + M_{yx} w_{,yx}\big) = \frac{D}{2}\big(w_{,xx}^2 + w_{,yy}^2 + 2w_{,xy}^2\big)\ . \tag{7.102}$$

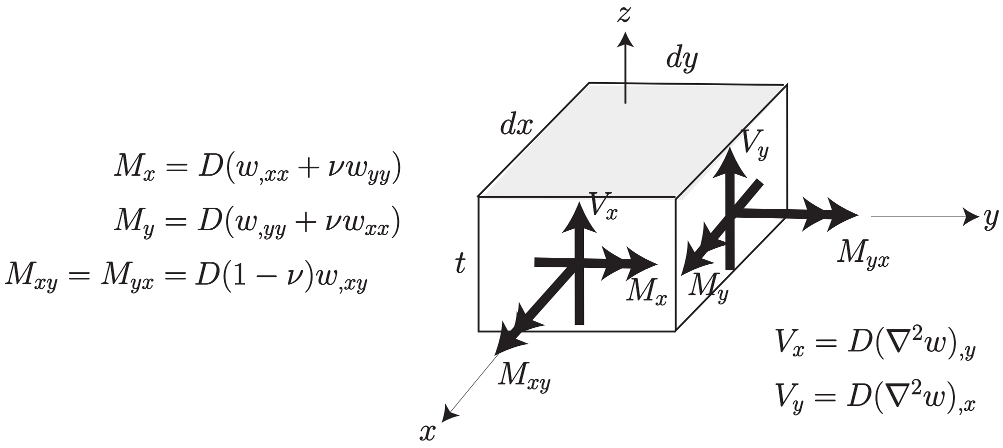

$M_x = D(w_{,xx} + \nu w_{yy})$
$M_y = D(w_{,yy} + \nu w_{xx})$
$M_{xy} = M_{yx} = D(1-\nu)w_{,xy}$
z
dy
dx
V_y
V_x
t
M_x
M_y
M_{yx}
y
M_{xy}
x
$V_x = D(\nabla^2 w)_{,y}$
$V_y = D(\nabla^2 w)_{,x}$

Figure 7.17: A differential element $dx \times dy \times t$ of a thin plate. A bold arrow represents a force and thus V_x and V_y are (shear) forces. A bold arrow with two arrow heads represents a moment whose sense is given by the right hand rule. Thus M_{xy} and M_{yx} are (twisting) moments while M_x and M_y are (bending) moments.

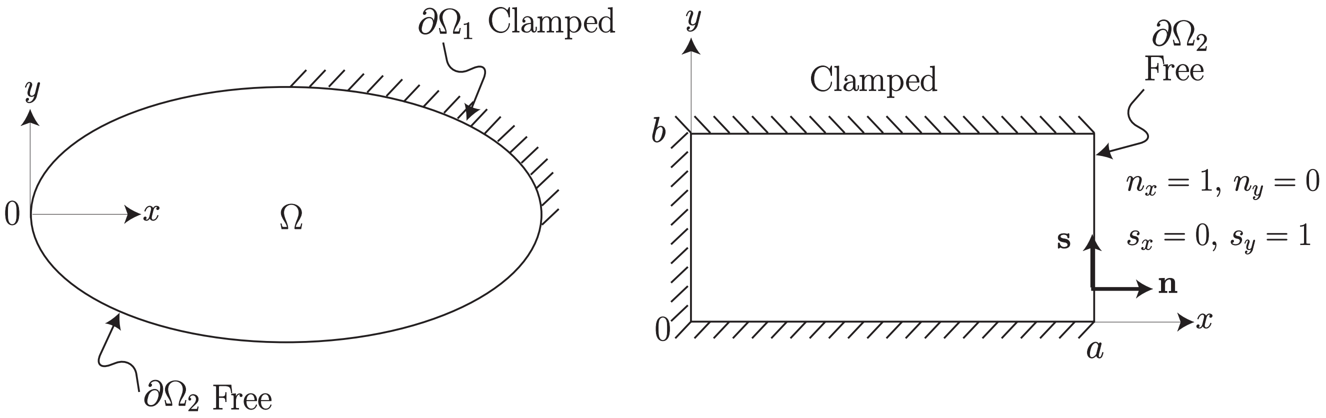


Figure 7.18: Left: A thin elastic plate whose mid-plane occupies a region Ω of the x, y-plane. The segment $\partial\Omega_1$ of the boundary is clamped while the remainder $\partial\Omega_2$ is free of loading. Right: A rectangular $a \times b$ plate with a load free edge at its right hand side $x = a, 0 \le y \le b$.

It is worth noting the following puzzling question: Consider the rectangular plate shown in the right hand diagram of Figure 7.18. Based on Figure 7.17 we know that there is a bending moment M_x, a twisting moment M_{xy}, and a shear force V_x acting on any surface $x = \mathsf{constant}$ in the plate. Therefore, in particular, since the right hand edge $x = a$ is free of loading one would expect to have the three conditions $M_x = M_{xy} = V_x = 0$ along that boundary. However we will find that the differential equation to be solved in the interior of the plate requires (and can only accommodate) *two* boundary conditions at any point on the edge. The question then arises as to what the correct boundary conditions on this edge should be. Our variational approach will give us precisely two natural boundary conditions on this edge. They will involve M_x, M_{xy} and V_x but will not require that each of them must vanish individually.

Consider a thin elastic plate whose mid-plane occupies a domain Ω of the x, y-plane as shown in the left hand diagram of Figure 7.18. A normal loading $p(x, y)$ is applied on the flat face of the plate in the $-z$-direction. A part of the plate boundary denoted by $\partial\Omega_1$ is clamped while the remainder $\partial\Omega_2$ is free of any external loading. Thus if $w(x, y)$ denotes the deflection of the plate in the z-direction we have the geometric boundary conditions

$$w = \partial w/\partial n = 0 \qquad \text{for } (x, y) \in \partial\Omega_1. \tag{7.103}$$

The total potential energy of the system is

$$\Phi\{w\} = \int_\Omega \left[\frac{D}{2}\Big(w_{,xx}^2 + 2w_{,xy}^2 + w_{,yy}^2\Big) - p\,w\right] dA \tag{7.104}$$

where the first group of terms represents the elastic energy in the plate and the last term represents the potential energy of the pressure loading (the negative sign arising from the fact that p acts in the minus z-direction while w is the deflection in the positive z direction). This functional Φ is defined on the set of all kinematically admissible deflection fields which is the set of all suitably smooth functions $w(x, y)$ that satisfy the geometric requirements (7.103). The actual deflection field is the one that minimizes the potential energy Φ over this set.

We now determine the Euler equation and natural boundary conditions associated with (7.104) by calculating the first variation of $\Phi\{w\}$ and setting it equal to zero:

$$\int_\Omega \Big(w_{,xx}\delta w_{,xx} + 2w_{,xy}\delta w_{,xy} + w_{,yy}\delta w_{,yy} - \frac{p}{D}\delta w\Big) dA = 0. \tag{7.105}$$

To simplify this we begin be rearranging the terms into a form that will allow us to use the divergence theorem, thereby converting part of the area integral on Ω into a boundary integral on $\partial\Omega$. In order to use the divergence theorem we must write the integrand so that it involves terms of the form $(\ldots)_{,x} + (\ldots)_{,y}$; see (7.96). Accordingly we rewrite (7.105) as

$$\begin{aligned}
0 &= \int_\Omega \Big(w_{,xx}\delta w_{,xx} + 2w_{,xy}\delta w_{,xy} + w_{,yy}\delta w_{,yy} - (p/D)\delta w\Big)\,dA,\\
&= \int_\Omega \Big[\Big(w_{,xx}\delta w_{,x} + w_{,xy}\delta w_{,y}\Big)_{,x} + \Big(w_{,xy}\delta w_{,x} + w_{,yy}\delta w_{,y}\Big)_{,y}\\
&\qquad -w_{,xxx}\delta w_{,x} - w_{,xxy}\delta w_{,y} - w_{,xyy}\delta w_{,x} - w_{,yyy}\delta w_{,y} - (p/D)\delta w\Big]\,dA,\\
&= \int_{\partial\Omega} \Big[\Big(w_{,xx}\delta w_{,x} + w_{,xy}\delta w_{,y}\Big)n_x + \Big(w_{,xy}\delta w_{,x} + w_{,yy}\delta w_{,y}\Big)n_y\Big]\,ds\\
&\qquad -\int_\Omega \Big[w_{,xxx}\delta w_{,x} + w_{,xxy}\delta w_{,y} + w_{,xyy}\delta w_{,x} + w_{,yyy}\delta w_{,y} + (p/D)\delta w\Big]\,dA,\\
&= \int_{\partial\Omega} I_1\,ds - \int_\Omega I_2\,dA.
\end{aligned} \tag{7.106}$$

We have used the divergence theorem (7.96) in going from the second equation above to the third equation. In order to facilitate further simplification, in the last step we have let I_1 and I_2 denote the integrands of the boundary and area integrals.

To simplify the area integral in (7.106) we again rearrange the terms in I_2 into a form that will allow us to use the divergence theorem. Thus

$$\begin{aligned}
\textstyle\int_\Omega I_2\,dA &= \int_\Omega \Big[w_{,xxx}\delta w_{,x} + w_{,xxy}\delta w_{,y} + w_{,xyy}\delta w_{,x} + w_{,yyy}\delta w_{,y} + p/D\,\delta w\Big]\,dA,\\
&= \int_\Omega \Big[\Big(w_{,xxx}\delta w + w_{,xyy}\delta w\Big)_{,x} + \Big(w_{,xxy}\delta w + w_{,yyy}\delta w\Big)_{,y}\\
&\qquad\qquad -\Big(w_{,xxxx} + 2w_{,xxyy} + w_{,yyyy} - p/D\Big)\delta w\Big]\,dA,\\
&= \int_{\partial\Omega} \Big[\Big(w_{,xxx}\delta w + w_{,xyy}\delta w\Big)n_x + \Big(w_{,xxy}\delta w + w_{,yyy}\delta w\Big)n_y\Big]\,ds\\
&\qquad\qquad -\int_\Omega \Big(\nabla^4 w - (p/D)\Big)\,\delta w\,dA,\\
&= \int_{\partial\Omega} \Big[w_{,xxx}n_x + w_{,xyy}n_x + w_{,xxy}n_y + w_{,yyy}n_y\Big]\,\delta w\,ds\\
&\qquad\qquad -\int_\Omega \Big(\nabla^4 w - p/D\Big)\,\delta w\,dA,\\
&= \int_{\partial\Omega} P_1\,\delta w\,ds - \int_\Omega P_2\,\delta w\,dA,
\end{aligned} \tag{7.107}$$

where we have set

$$P_1 = w_{,xxx}n_x + w_{,xyy}n_x + w_{,xxy}n_y + w_{,yyy}n_y \qquad \text{and} \qquad P_2 = \nabla^4 w - p/D. \tag{7.108}$$

In the preceding calculation, we have used the divergence theorem (7.96) in going from the second equation in (7.107) to the third equation, and we have set

$$\nabla^4 w = \nabla^2(\nabla^2 w) = w_{,xxxx} + 2w_{,xxyy} + w_{,yyyy}.$$

Next we simplify the boundary term in (7.106) by converting the derivatives of the variation with respect to x and y into derivatives with respect to normal and tangential coordinates n and s. To do this we use the fact that $\boldsymbol{\nabla}\delta w = \delta w_{,x}\mathbf{i} + \delta w_{,y}\mathbf{j} = \delta w_{,n}\mathbf{n} + \delta w_{,s}\mathbf{s}$ from which it follows that $\delta w_{,x} = \delta w_{,n}\, n_x + \delta w_{,s}\, s_x$ and $\delta w_{,y} = \delta w_{,n}\, n_y + \delta w_{,s}\, s_y$. Thus from (7.106),

$$\begin{aligned}
\textstyle\int_{\partial\Omega} I_1\, ds \; &= \int_{\partial\Omega} \Big[\Big(w_{,xx}n_x\delta w_{,x} + w_{,xy}n_x\delta w_{,y}\Big) + \Big(w_{,xy}n_y\delta w_{,x} + w_{,yy}n_y\delta w_{,y}\Big)\Big]\, ds,\\
&= \int_{\partial\Omega} \Big[\Big(w_{,xx}n_x + w_{,xy}n_y\Big)\delta w_{,x} + \Big(w_{,xy}n_x + w_{,yy}n_y\Big)\delta w_{,y}\Big]\, ds,\\
&= \int_{\partial\Omega} \Big[\Big(w_{,xx}n_x + w_{,xy}n_y\Big)\Big(\delta w_{,n}n_x + \delta w_{,s}s_x\Big) + \Big(w_{,xy}n_x + w_{,yy}n_y\Big)\Big(\delta w_{,n}n_y + \delta w_{,s}s_y\Big)\Big]\, ds,\\
&= \int_{\partial\Omega} \Big[\Big(w_{,xx}n_x^2 + w_{,xy}n_xn_y + w_{,xy}n_xn_y + w_{,yy}n_y^2\Big)\delta w_{,n}\\
&\qquad\qquad + \Big(w_{,xx}n_xs_x + w_{,xy}s_xn_y + w_{,xy}n_xs_y + w_{,yy}n_ys_y\Big)\delta w_{,s}\Big]\, ds,\\
&= \int_{\partial\Omega} \Big[\Big(w_{,xx}n_x^2 + w_{,xy}n_xn_y + w_{,xy}n_xn_y + w_{,yy}n_y^2\Big)\delta w_{,n} \; + \; I_3\Big]\, ds.
\end{aligned} \tag{7.109}$$

To further simplify this we have set I_3 equal to the last expression in (7.109) and this term can be written as

$$\begin{aligned}
\textstyle\int_{\partial\Omega} I_3\, ds \; &= \; \int_{\partial\Omega} \Big[\Big(w_{,xx}n_xs_x + w_{,xy}s_xn_y + w_{,xy}n_xs_y + w_{,yy}n_ys_y\Big)\, \delta w_{,s}\Big]\, ds,\\
&= \; \int_{\partial\Omega} \Big[\Big(w_{,xx}n_xs_x + w_{,xy}s_xn_y + w_{,xy}n_xs_y + w_{,yy}n_ys_y\Big)\, \delta w\Big]_{,s}\\
&\qquad\qquad - \Big[\Big(w_{,xx}n_xs_x + w_{,xy}s_xn_y + w_{,xy}n_xs_y + w_{,yy}n_ys_y\Big)_{,s}\, \delta w\Big]\, ds.
\end{aligned} \tag{7.110}$$

If a field $f(x, y)$ varies smoothly along $\partial\Omega$, and if the curve $\partial\Omega$ itself is smooth, then

$$\int_{\partial\Omega} \frac{\partial f}{\partial s}\, ds \quad = \; 0 \tag{7.111}$$

since this is an integral over a closed curve[8]. It follows from this that the first term in the last expression of (7.110) vanishes and so

$$\textstyle\int_{\partial\Omega} I_3\,ds \;=\; -\int_{\partial\Omega}\Big[\Big(w_{,xx}n_xs_x + w_{,xy}s_xn_y + w_{,xy}n_xs_y + w_{,yy}n_ys_y\Big)_{,s}\,\delta w\Big]\,ds. \tag{7.112}$$

Substituting (7.112) into (7.109) yields

$$\int_{\partial\Omega} I_1\,ds \;=\int_{\partial\Omega} P_3\,\frac{\partial}{\partial n}(\delta w)\,ds - \int_{\partial\Omega}\frac{\partial}{\partial s}(P_4)\,\delta w\,ds, \tag{7.113}$$

where we have set

$$\begin{aligned} &P_3 = w_{,xx}n_x^2 + w_{,xy}n_xn_y + w_{,xy}n_xn_y + w_{,yy}n_y^2,\\ &P_4 = w_{,xx}n_xs_x + w_{,xy}n_ys_x + w_{,xy}n_xs_y + w_{,yy}n_ys_y. \end{aligned} \tag{7.114}$$

Finally, substituting (7.113) and (7.107) into (7.106) leads to

$$\int_{\Omega} P_2\,\delta w\,dA \;-\; \int_{\partial\Omega}\left(P_1 + \frac{\partial}{\partial s}(P_4)\right)\,\delta w\,ds + \int_{\partial\Omega} P_3\,\frac{\partial}{\partial n}(\delta w)\,ds = 0 \tag{7.115}$$

which must hold for all admissible variations δw. First restrict attention to variations which vanish on the boundary $\partial\Omega$ and whose normal derivative $\partial(\delta w)/\partial n$ also vanish on $\partial\Omega$. This leads us to the Euler equation $P_2 = 0$:

$$\nabla^4 w - p/D = 0 \qquad \text{for} \quad (x,y)\in\Omega. \tag{7.116}$$

Returning to (7.115) with this gives

$$-\int_{\partial\Omega}\left(P_1 + \frac{\partial}{\partial s}(P_4)\right)\,\delta w\,ds + \int_{\partial\Omega} P_3\,\frac{\partial}{\partial n}(\delta w)\,ds = 0. \tag{7.117}$$

Since the portion $\partial\Omega_1$ of the boundary is clamped we have $w = \partial w/\partial n = 0$ for $(x,y)\in\partial\Omega_1$. Thus the variations δw and $\partial(\delta w)/\partial n$ must also vanish on $\partial\Omega_1$. Thus (7.117) simplifies to

$$-\int_{\partial\Omega_2}\left(P_1 + \frac{\partial}{\partial s}(P_4)\right)\,\delta w\,ds + \int_{\partial\Omega_2} P_3\,\frac{\partial}{\partial n}(\delta w)\,ds = 0 \tag{7.118}$$

for variations δw and $\partial(\delta w)/\partial n$ that are arbitrary on $\partial\Omega_2$ where $\partial\Omega_2$ is the complement of $\partial\Omega_1$, i.e. $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2$. Thus we conclude that $P_1 + \partial P_4/\partial s = 0$ and $P_3 = 0$ on $\partial\Omega_2$:

$$\left.\begin{aligned} &w_{,xxx}n_x + w_{,xyy}n_x + w_{,xxy}n_y + w_{,yyy}n_y\\ &\qquad\qquad + \tfrac{\partial}{\partial s}(w_{,xx}n_xs_x + w_{,xy}n_ys_x + w_{,xy}n_xs_y + w_{,yy}n_ys_y) = 0\\ &w_{,xx}n_x^2 + w_{,xy}n_xn_y + w_{,xy}n_xn_y + w_{,yy}n_y^2 \;=\; 0, \end{aligned}\right\}\text{for } (x,y)\in\partial\Omega_2. \tag{7.119}$$

[8] In the present setting one would have this degree of smoothness if there are no concentrated loads applied on the boundary of the plate $\partial\Omega$ and the boundary curve itself has no corners.

Thus, in summary, the *Kirchhoff theory of plates* for the problem at hand requires that one solve the field equation (7.116) on Ω subjected to the displacement boundary conditions (7.103) on $\partial\Omega_1$ and the natural boundary conditions (7.119) on $\partial\Omega_2$.

Remark: If we define the moments M_n, M_{ns} and force V_n by

$$\begin{aligned} M_n &= -D\left(w_{,xx}\, n_x n_x + w_{,xy}\, n_x n_y + w_{,yx}\, n_y n_x + w_{,yy}\, n_y n_y\right) \\ M_{ns} &= -D\left(w_{,xx}\, n_x\, s_x + w_{,xy}\, n_y\, s_x + w_{,yx}\, n_x\, s_y + w_{,yy}\, n_y\, s_y\right) \\ V_n &= -D\left(w_{,xxx}\, n_x + w_{,xyy}\, n_x + w_{,yxx}\, n_y + w_{,yyy}\, n_y\right) \end{aligned} \tag{7.120}$$

then the two natural boundary conditions can be written as

$$M_n = 0, \qquad \frac{\partial}{\partial s}\big(M_{ns}\big) + V_n = 0. \tag{7.121}$$

As a special case suppose that the plate is rectangular, $0 \leq x \leq a, 0 \leq y \leq b$ and that the right edge $x = a, 0 \leq y \leq b$ is free of load; see the right diagram in Figure 7.18. Then $n_x = 1, n_y = 0, s_x = 0, s_y = 1$ on $\partial\Omega_2$ and so (7.120) simplifies to

$$\begin{aligned} M_n &= -D\, w_{,xx} \\ M_{ns} &= -D\, w_{,yx} \\ V_n &= -D\left(w_{,xxx} + w_{,xyy}\right) \end{aligned} \tag{7.122}$$

which because of (7.100) shows that in this case $M_n = M_x, M_{ns} = M_{xy}, V_n = V_x$. Thus the natural boundary conditions (7.121) can be written as

$$M_x = 0, \qquad \frac{\partial}{\partial y}\big(M_{xy}\big) + V_x = 0. \tag{7.123}$$

This answers the question we posed soon after (7.101) as to what the correct boundary conditions on a free edge should be. We had noted that intuitively we would have expected the moments and forces to vanish on a free edge and therefore that $M_x = M_{xy} = V_x = 0$ there; but this is in contradiction to the mathematical fact that the differential equation (7.116) only requires two conditions at each point on the boundary. The two natural boundary conditions (7.123) require that certain combinations of M_x, M_{xy}, V_x vanish but not that all three vanish.

Example 3: Minimal surface equation. Let $\mathcal{C}$ be a closed curve in $\mathbb{R}^3$ as sketched in Figure 7.19. From among all surfaces $\mathcal{S}$ in $\mathbb{R}^3$ that have $\mathcal{C}$ as its boundary, we wish to determine the surface that has minimum area. As a physical example, if $\mathcal{C}$ corresponds to a wire loop which we dip in a soapy solution, a thin soap film will form across $\mathcal{C}$. The surface that forms is the one that, from among all possible surfaces $\mathcal{S}$ that are bounded by $\mathcal{C}$, minimizes the total surface energy of the film; which (if the surface energy density is constant) is the surface with minimal area.

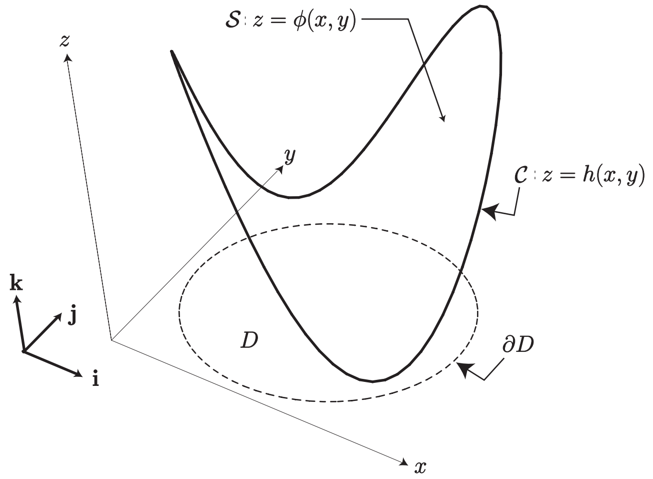


Figure 7.19: The closed curve $\mathcal{C}$ in $\mathbb{R}^3$ is given. From among all surfaces $\mathcal{S}$ in $\mathbb{R}^3$ that have $\mathcal{C}$ as its boundary, the surface with minimal area is to be sought. The curve ∂D is the projection of $\mathcal{C}$ onto the x, y-plane.

Let $\mathcal{C}$ be a closed curve in $\mathbb{R}^3$. Suppose that its projection onto the x, y-plane is denoted by ∂D and let D denote the simply connected region contained within ∂D; see Figure 7.19. Suppose that $\mathcal{C}$ is characterized by $z = h(x,y)$ for $(x,y) \in \partial D$. Let $z = \phi(x,y)$ for $(x,y) \in D$ describe a surface $\mathcal{S}$ in $\mathbb{R}^3$ that has $\mathcal{C}$ as its boundary; necessarily $\phi = h$ on ∂D. Thus the admissible set of functions we are considering are

$$\mathsf{A}\{\phi \,|\, \phi : D \to \mathbb{R},\ \phi \in C^1(D),\ \phi = h \text{ on } \partial D\}\ .$$

Consider a rectangular differential element on the x, y-plane that is contained within D. The vector joining (x,y) to $(x+dx, y)$ is $\mathbf{dx} = dx\,\mathbf{i}$ while the vector joining (x,y) to $(x, y+dy)$ is $\mathbf{dy} = dy\,\mathbf{j}$. If $\mathbf{du}$ and $\mathbf{dv}$ are vectors on the surface $z = \phi(x,y)$ whose projections are $\mathbf{dx}$ and $\mathbf{dy}$ respectively, then we know that

$$\mathbf{du} = dx\,\mathbf{i} + \phi_x\,dx\,\mathbf{k}, \qquad \mathbf{dv} = dy\,\mathbf{j} + \phi_y\,dy\,\mathbf{k}.$$

The vectors $\mathbf{du}$ and $\mathbf{dv}$ define a parallelogram on the surface $z = \phi(x,y)$ and the area of this parallelogram is $|\mathbf{du} \times \mathbf{dv}|$. Thus the area of a differential element on $\mathcal{S}$ is

$$|\mathbf{du} \times \mathbf{dv}| = \Big| - \phi_x dxdy\,\mathbf{i} - \phi_y dxdy\,\mathbf{j} + dxdy\,\mathbf{k}\Big| = \sqrt{1 + \phi_x^2 + \phi_y^2}\,dxdy.$$

Consequently the problem at hand is to minimize the functional

$$F\{\phi\} = \int_D \sqrt{1+\phi_x^2+\phi_y^2}\; dA.$$

over the set of admissible functions

$$\mathsf{A} = \{\phi \mid \phi : D \to \mathbb{R},\, \phi \in C^2(D),\, \phi = h \text{ on } \partial D\}.$$

It is left as an exercise to show that setting the first variation of F equal to zero leads to the so-called minimal surface equation

$$(1+\phi_y^2)\phi_{xx} - 2\phi_x\phi_y\phi_{xy} + (1+\phi_x^2)\phi_{yy} = 0.$$

Remark: See en.wikipedia.org/wiki/Soap_bubble and www.susqu.edu/facstaff/b/brakke/ for additional discussion.

7.9 Second variation. Another necessary condition for a minimum.

In order to illustrate the basic ideas of this section in the simplest possible setting, we confine the discussion to the particular functional

$$F\{\phi\} = \int_0^1 f(x,\phi,\phi')dx$$

defined over a set of admissible functions A. Suppose that a particular function ϕ minimizes F, and that for some given function η, the one-parameter family of functions $\phi+\varepsilon\eta$ are admissible for all sufficiently small values of the parameter ε. Define $\hat{F}(\varepsilon)$ by

$$\hat{F}(\varepsilon) = F\{\phi+\varepsilon\eta\} = \int_0^1 f(x,\phi+\varepsilon\eta,\phi'+\varepsilon\eta')dx,$$

so that by Taylor expansion,

$$\hat{F}(\varepsilon) = \hat{F}(0) + \varepsilon\hat{F}'(0) + \frac{\varepsilon^2}{2}\hat{F}''(0) + O(\varepsilon^3),$$

where

$$\begin{aligned}
\hat{F}(0) &= \textstyle\int_0^1 f(x,\phi,\phi')dx = F\{\phi\},\\
\varepsilon\hat{F}'(0) &= \delta F\{\phi,\eta\},\\
\varepsilon^2 F''(0) &= \varepsilon^2 \textstyle\int_0^1 \{f_{\phi\phi}\eta^2 + 2f_{\phi\phi'}\eta\eta' + f_{\phi'\phi'}(\eta')^2\}\,dx \overset{\text{def}}{=} \delta^2 F\{\phi,\eta\}.
\end{aligned}$$

Since ϕ minimizes F, it follows that $\varepsilon = 0$ minimizes $\hat{F}(\varepsilon)$, and consequently that

$$\delta^2 F\{\phi, \eta\} \geq 0,$$

in addition to the requirement $\delta F\{\phi, \eta\} = 0$. Thus a necessary condition for a function ϕ to minimize a functional F is that the second variation of F be non-negative for all admissible variations $\delta\phi$:

$$\delta^2 F\{\phi, \delta\phi\} = \int_0^1 \left\{ f_{\phi\phi}(\delta\phi)^2 + 2f_{\phi\phi'}(\delta\phi)(\delta\phi') + f_{\phi'\phi'}(\delta\phi')^2 \right\} dx \geq 0, \tag{7.124}$$

where we have set $\delta\phi = \varepsilon\eta$. The inequality is reversed if ϕ maximizes F.

The condition $\delta^2 F\{\phi, \eta\} \geq 0$ is necessary but not sufficient for the functional F to have a minimum at ϕ. We shall not discuss sufficient conditions in general in these notes.

Proposition: Legendre Condition: A necessary condition for (7.124) to hold is that

$$f_{\phi'\phi'}(x, \phi(x), \phi'(x)) \geq 0 \qquad \text{for} \quad 0 \leq x \leq 1$$

for the minimizing function ϕ.

Example: Consider a curve in the x, y-plane characterized by $y = \phi(x)$ that begins at $(0, \phi_0)$ and ends at $(1, \phi_1)$. From among all such curves, find the one that, when rotated about the x-axis, generates the surface of minimum area.

Thus we are asked to minimize the functional

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi')\, dx \qquad \text{where} \quad f(x, \phi, \phi') = \phi\sqrt{1 + (\phi')^2},$$

over a set of admissible functions that satisfy the boundary conditions $\phi(0) = \phi_0, \phi(1) = \phi_1$.

A function ϕ that minimizes F must satisfy the boundary value problem consisting of the Euler equation and the given boundary conditions:

$$\left.\begin{array}{c} \dfrac{d}{dx}\left(\dfrac{\phi\phi'}{\sqrt{1 + (\phi')^2}} \right) - \sqrt{1 + (\phi')^2} = 0, \\ \phi(0) = \phi_0, \qquad \phi(1) = \phi_1. \end{array}\right\}$$

The general solution of this Euler equation is

$$\phi(x) = \alpha \cosh \frac{x - \beta}{\alpha} \qquad \text{for} \quad 0 \leq x \leq 1,$$

where the constants α and β are determined through the boundary conditions. To test the Legendre condition we calculate $f_{\phi'\phi'}$ and find that

$$f_{\phi'\phi'} = \frac{\phi}{(\ \sqrt{1+(\phi')^2}\ )^3},$$

which, when evaluated at the particular function $\phi(x) = \alpha \cosh(x-\beta)/\alpha$ yields

$$f_{\phi'\phi'}\big|_{\phi=\alpha\cosh(x-\beta)/\alpha} = \frac{\alpha}{\cosh^2(x-\beta)/\alpha}.$$

Therefore as long as $\alpha > 0$ the Legendre condition is satisfied.

7.10 Sufficient condition for minimization of convex functionals

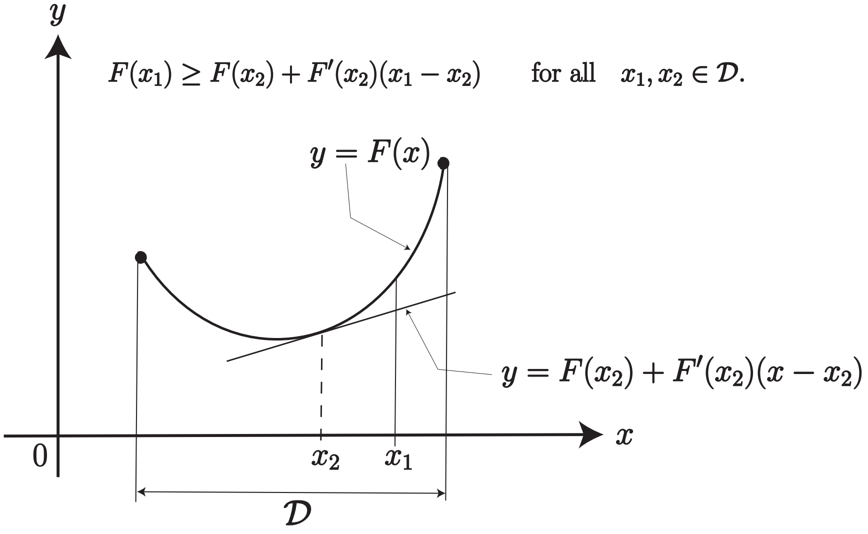

y
$F(x_1) \geq F(x_2) + F'(x_2)(x_1 - x_2)$ for all $x_1, x_2 \in \mathcal{D}$.
$y = F(x)$
$y = F(x_2) + F'(x_2)(x - x_2)$
x
0
x_2
x_1
$\mathcal{D}$

Figure 7.20: A convex curve $y = F(x)$ lies above the tangent line through any point of the curve.

We now turn to a brief discussion of sufficient conditions for a minimum for a special class of functionals. It is useful to begin by reviewing the question of finding the minimum of a real-valued function of a real variable. A function $F(x)$ defined for $x \in \mathsf{A}$ with continuous first derivatives is said to be *convex* if

$$F(x_1) \geq F(x_2) + F'(x_2)(x_1 - x_2) \qquad \text{for all} \quad x_1, x_2 \in \mathsf{A};$$

see Figure 7.20 for a geometric interpretation of convexity. If a convex function has a stationary point, say at x_o, then it follows by setting $x_2 = x_o$ in the preceding equation that x_o is a minimizer of F. Therefore a stationary point of a convex function is necessarily a minimizer. If F is strictly convex on A, i.e. if F is convex and $F(x_1) = F(x_2) + F'(x_2)(x_1 - x_2)$ if and only if $x_1 = x_2$, then F can only have one stationary point and so can only have one interior minimum.

This is also true for a real-valued function F with continuous first derivatives on a domain A in $\mathbb{R}^n$, where *convexity* is defined by[9]

$$F(\mathbf{x}_1) \geq F(\mathbf{x}_2) + \delta F(\mathbf{x}_2, \mathbf{x}_1 - \mathbf{x}_2) \qquad \text{for all} \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathsf{A}.$$

If a convex function has a stationary point at, say, $\mathbf{x}_o$, then since $\delta F(\mathbf{x}_o, \mathbf{y}) = 0$ for all $\mathbf{y}$ it follows that $\mathbf{x}_o$ is a minimizer of F. Therefore a stationary point of a convex function is necessarily a minimizer. If F is strictly convex on A, i.e. if F is convex and $F(\mathbf{x}_1) = F(\mathbf{x}_2) + \delta F(\mathbf{x}_2, \mathbf{x}_1 - \mathbf{x}_2)$ if and only if $\mathbf{x}_1 = \mathbf{x}_2$, then F can have only one stationary point and so can have only one interior minimum.

We now turn to a functional $F\{\phi\}$ which is said to be *convex* on A if

$$F\{\phi + \eta\} \geq F\{\phi\} + \delta F\{\phi, \eta\} \qquad \text{for all} \quad \phi, \phi + \eta \in \mathsf{A}.$$

If F is stationary at $\phi_o \in \mathsf{A}$, then $\delta F\{\phi_o, \eta\} = 0$ for all admissible η, and it follows that ϕ_o is in fact a minimizer of F. Therefore a stationary point of a convex functional is necessarily a minimizer.

For example, consider the generic functional

$$F\{\phi\} = \int_0^1 f(x, \phi, \phi')dx. \tag{7.125}$$

Then

$$\delta F\{\phi, \eta\} = \int_0^1 \left(\frac{\partial f}{\partial \phi}\eta + \frac{\partial f}{\partial \phi'}\eta' \right) dx$$

and so the convexity condition $F\{\phi + \eta\} - F\{\phi\} \geq \delta F\{\phi, \eta\}$ takes the special form

$$\int_0^1 \Big[f(x, \phi + \eta, \phi' + \eta') - f(x, \phi, \phi') \Big] dx \geq \int_0^1 \left(\frac{\partial f}{\partial \phi}\eta + \frac{\partial f}{\partial \phi'}\eta' \right) dx. \tag{7.126}$$

In general it might not be simple to test whether this condition holds in a particular case. It is readily seen that a sufficient condition for (7.126) to hold is that the integrands satisfy

[9] See equation (**??**) for the definition of $\delta F(\mathbf{x}, \mathbf{y})$.

the inequality

$$f(x, y+v, z+w) - f(x,y,z) \geq \frac{\partial f}{\partial y} v + \frac{\partial f}{\partial z} w \tag{7.127}$$

for all $(x, y, z), (x, y+v, z+w)$ in the domain of f. This is precisely the requirement that the function $f(x, y, z)$ be a convex function of y, z at fixed x.

Thus in summary: if the integrand f of the functional F defined in (7.125) satisfies the convexity condition (7.127), then, a function ϕ that extremizes F is in fact a minimizer of F. Note that this is simply a sufficient condition for ensuring that an extremum is a minimum.

Remark: In the special case where $f(x, y, z)$ is independent of y, one sees from basic calculus that if $\partial^2 f/\partial z^2 > 0$ then $f(x, z)$ is a strictly convex function of z at each fixed x.

Example: Geodesics. Find the curve of shortest length that lies entirely on a circular cylinder of radius a, beginning (in circular cylindrical coordinates (r, θ, ξ)) at (a, θ_1, ξ_1) and ending at (a, θ_2, ξ_2) as shown in the figure.

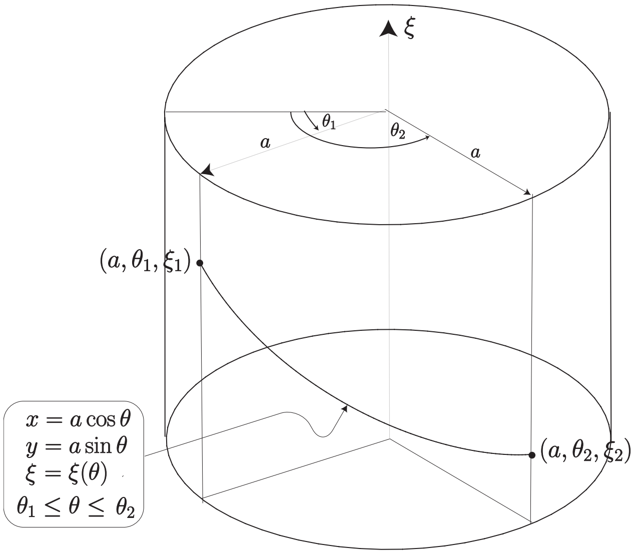


Figure 7.21: A curve that lies entirely on a circular cylinder of radius a, beginning (in circular cylindrical coordinates) at (a, θ_1, ξ_1) and ending at (a, θ_2, ξ_2).

We can characterize a curve in $\mathbb{R}^3$ using a parametric characterization using circular cylindrical coordinates by $r = r(\theta), \xi = \xi(\theta)$, $\theta_1 \leq \theta \leq \theta_2$. When the curve lies on the surface of a circular cylinder of radius a this specializes to

$$r = a, \qquad \xi = \xi(\theta) \qquad \text{for} \quad \theta_1 \leq \theta \leq \theta_2.$$

Since the arc length can be written as

$$ds \;=\; \sqrt{dr^2 + r^2 d\theta^2 + d\xi^2} \;=\; \sqrt{\big(r'(\theta)\big)^2 + \big(r(\theta)\big)^2 + \big(\xi'(\theta)\big)^2}\, d\theta \;==\; \sqrt{\big(a^2 + \big(\xi'(\theta)\big)^2}\, d\theta.$$

our task is to minimize the functional

$$F\{\xi\} = \int_{\theta_1}^{\theta_2} f(\theta, \xi(\theta), \xi'(\theta))\, d\theta \qquad \text{where} \quad f(x,y,z) = \sqrt{a^2 + z^2}$$

over the set of all suitably smooth functions $\xi(\theta)$ defined for $\theta_1 \leq \theta \leq \theta_2$ which satisfy $\xi(\theta_1) = \xi_1$, $\xi(\theta_2) = \xi_2$.

Evaluating the necessary condition $\delta F = 0$ leads to the Euler equation. This second order differential equation for $\xi(\theta)$ can be readily solved, which after using the boundary conditions $\xi(\theta_1) = \xi_1$, $\xi(\theta_2) = \xi_2$ leads to

$$\xi(\theta) = \xi_1 + \left(\frac{\xi_1 - \xi_2}{\theta_1 - \theta_2}\right)(\theta - \theta_1). \tag{7.128}$$

Direct differentiation of $f(x,y,z) = \sqrt{a^2+z^2}$ shows that

$$\frac{\partial^2 f}{\partial z^2} = \frac{a^2}{(a^2+z^2)^{3/2}} \;> 0$$

and so f is a strictly convex function of z. Thus the curve of minimum length is given uniquely by (7.128) – a helix. Note that if the circular cylindrical surface is cut along a vertical line and unrolled into a flat sheet, this curve unfolds into a straight line.

7.11 Direct method of the calculus of variations and minimizing sequences.

We now turn to a different method of seeking minima, and for purposes of introduction, begin by reviewing the familiar case of a real-valued function $f(x)$ of a real variable $x \in (-\infty, \infty)$. Consider the specific example $f(x) = x^2$. This function is nonnegative and has a minimum value of zero which it attains at $x = 0$. Consider the sequence of numbers

$$x_0,\; x_1,\; x_2,\; x_3 \ldots x_k, \ldots \qquad \text{where} \quad x_k = \frac{1}{2^k}\;,$$

and note that

$$\lim_{k\to\infty} f(x_k) = 0.$$

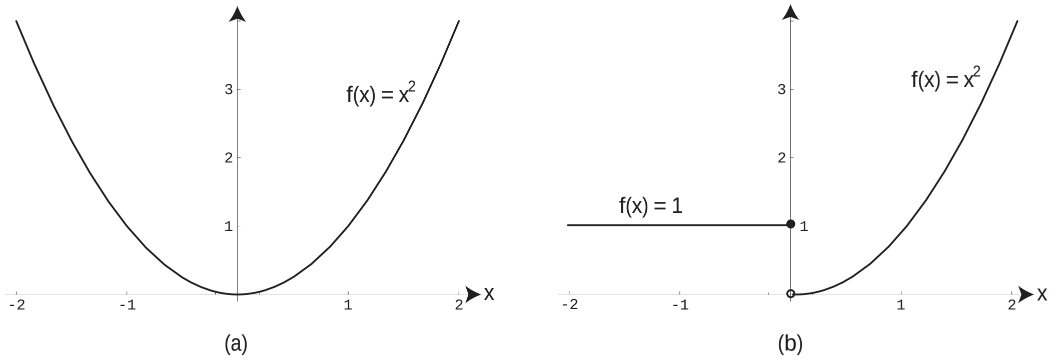


Figure 7.22: (a) The function $f(x) = x^2$ for $-\infty < x < \infty$ and (b) the function $f(x) = 1$ for $x \leq 0$, $f(x) = x^2$ for $x > 0$.

The sequence $1/2,\ 1/2^2,\ \ldots,\ 1/2^k,\ \ldots$ is called a minimizing sequence in the sense that the value of the function $f(x_k)$ converges to the minimum value of f as $k \to \infty$. Moreover, observe that

$$\lim_{k\to\infty} x_k = 0$$

as well, and so the sequence itself converges to the minimizer of f, i.e. to $x = 0$. This latter feature is true because in this example

$$f(\lim_{k\to\infty} x_k) = \lim_{n\to\infty} f(x_k).$$

As we know, not all functions have a minimum value, even if they happen to have a finite greatest lower bound. We now consider an example to illustrate the fact that a minimizing sequence can be used to find the greatest lower bound of a function that does not have a minimum. Consider the function $f(x) = 1$ for $x \leq 0$ and $f(x) = x^2$ for $x > 0$. This function is non-negative, and in fact, it can take values arbitrarily close to the value 0. However it does *not* have a minimum value since there is no value of x for which $f(x) = 0$; (note that $f(0) = 1$). The greatest lower bound or infimum (denoted by "inf") of f is

$$\inf_{-\infty<x<\infty} f(x) = 0.$$

Again consider the sequence of numbers

$$x_0,\ x_1,\ x_2,\ x_3 \ldots x_k, \ldots \qquad \text{where} \quad x_k = \frac{1}{2^k},$$

and note that

$$\lim_{k\to\infty} f(x_k) = 0.$$

In this case the value of the function $f(x_k)$ converges to the infimum of f as $k \to \infty$. However since

$$\lim_{k\to\infty} x_k = 0$$

the limit of the sequence itself is $x = 0$ and $f(0)$ is not the infimum of f. This is because in this example

$$f(\lim_{k\to\infty} x_k) \neq \lim_{n\to\infty} f(x_k).$$

Returning now to a functional, suppose that we are to find the infimum (or the minimum if it exists) of a functional $F\{\phi\}$ over an admissible set of functions A. Let

$$\inf_{\phi\in\mathsf{A}} F\{\phi\} = m \ (> -\infty).$$

Necessarily there must exist a sequence of functions $\phi_1, \phi_2, \ldots$ in A such that

$$\lim_{n\to\infty} F\{\phi_k\} = m;$$

such a sequence is called a minimizing sequence.

If the sequence $\phi_1, \phi_2, \ldots$ converges to a limiting function ϕ_*, and *if*

$$F\{\lim_{n\to\infty} \phi_k\} = \lim_{n\to\infty} F\{\phi_k\},$$

then it follows that $F\{\phi_*\} = m$ and the function ϕ_* is the minimizer of F. The functions ϕ_k of a minimizing sequence can be considered to be approximate solutions of the minimization problem.

Just as in the second example of this section, in some variational problems the limiting function ϕ_* of a minimizing sequence $\phi_1, \phi_2, \ldots$ does *not* minimize the functional F; see the last Example of this section.

7.11.1 The Ritz method

Suppose that we are to minimize a functional $F\{\phi\}$ over an admissible set A. Consider an infinite sequence of functions $\phi_1, \phi_2, \ldots$ in A. Let A_p be the subset of functions in A that can be expressed as a linear combination of the first p functions $\phi_1, \phi_2, \ldots \phi_p$. In order to minimize F over the subset A_p we must simply minimize

$$\widehat{F}(\alpha_1, \alpha_2, \ldots, \alpha_p) = F\{\alpha_1\phi_1 + \alpha_2\phi_2 + \ldots + \alpha_p\phi_p\}$$

with respect to the real parameters $\alpha_1, \alpha_2, \ldots \alpha_p$. Suppose that the minimum of F on A_p is denoted by m_p. Clearly $\mathsf{A}_1 \subset \mathsf{A}_2 \subset \mathsf{A}_3 \ldots \subset \mathsf{A}$ and therefore $m_1 \geq m_2 \geq m_3 \ldots$[10]. Thus, in the so-called *Ritz Method*, we minimize F over a subset A_p to find an approximate minimizer; moreover, increasing the value of p improves the approximation in the sense of the preceding footnote.

Example: Consider an elastic bar of length L and modulus E that is fixed at both ends and carries a distributed axial load $b(x)$. A displacement field $u(x)$ must satisfy the boundary conditions $u(0) = u(L) = 0$ and the associated potential energy is

$$F\{u\} = \int_0^L \frac{1}{2}E(u')^2 dx - \int_0^L bu \; dx.$$

We now use the Ritz method to find an approximate displacement field that minimizes F. Consider the sequence of functions $v_1, v_2, v_3, \ldots$ where

$$v_p = \sin \frac{p\pi x}{L};$$

observe that $v_p(0) = v_p(L) = 0$ for all intergers p. Consider the function

$$u_n(x) = \sum_{p=1}^{n} \alpha_p \sin \frac{p\pi x}{L}$$

for any integer $n \geq 1$ and evaluate

$$\widehat{F}(\alpha_1, \alpha_2, \ldots \alpha_n) = F\{u_n\} = \int_0^L \frac{1}{2}E(u_n')^2 dx - \int_0^L bu_n \; dx.$$

Since

$$\int_0^L 2\cos\frac{p\pi x}{L}\cos\frac{q\pi x}{L} dx = \begin{cases} 0 & \text{for} \quad p \neq q, \\ L & \text{for} \quad p = q, \end{cases}$$

it follows that

$$\int_0^L (u_n')^2 dx = \int_0^L \left(\sum_{p=1}^{n} \alpha_p \frac{p\pi}{L} \; \cos\frac{p\pi x}{L}\right)\left(\sum_{q=1}^{n} \alpha_q \frac{q\pi}{L} \; \cos\frac{q\pi x}{L}\right) dx = \frac{1}{2} \; \sum_{p=1}^{n} \alpha_p^2 \; \frac{p^2\pi^2}{L}$$

Therefore

$$\widehat{F}(\alpha_1, \alpha_2, \ldots \alpha_n) = F\{u_n\} = \sum_{p=1}^{n}\left(\frac{1}{4}E \; \alpha_p^2 \; \frac{p^2\pi^2}{L} - \alpha_p \int_0^L b \; \sin\frac{p\pi x}{L} \; dx\right) \tag{7.129}$$

[10] If the sequence $\phi_1, \phi_2, \ldots$ is complete, and the functional $F\{\phi\}$ is continuous in the appropriate norm, then one can show that $\lim_{p\to\infty} m_p = m$.

To minimize $\widehat{F}(\alpha_1, \alpha_2, \dots \alpha_n)$ with respect to α_p we set $\partial\widehat{F}/\partial\alpha_p = 0$. This leads to

$$\alpha_p = \frac{\int_0^L b \ \sin\frac{p\pi x}{L} \ dx}{E \ \frac{p^2\pi^2}{2L}} \qquad \text{for} \quad p = 1, 2, \dots n. \tag{7.130}$$

Therefore by substituting (7.130) into (7.129) we find that the n-term Ritz approximation of the energy is

$$-\sum_{p=1}^{n} \frac{1}{4} E \ \alpha_p^2 \ \frac{p^2\pi^2}{L} \qquad \text{where} \quad \alpha_p = \frac{\int_0^L b \ \sin\frac{p\pi x}{L} \ dx}{E \ \frac{p^2\pi^2}{2L}},$$

and the corresponding approximate displacement field is given by

$$u_n = \sum_{p=1}^{n} \alpha_p \sin\frac{p\pi x}{L} \qquad \text{where} \quad \alpha_p = \frac{\int_0^L b \ \sin\frac{p\pi x}{L} \ dx}{E \ \frac{p^2\pi^2}{2L}}.$$

References

1. J.L. Troutman, *Variational Calculus with Elementary Convexity*, Springer-Verlag, 1983.

2. C. Fox, *An Introduction to the Calculus of Variations*, Dover, 1987.

3. G. Bliss, Lectures on the Calculus of Variations, University of Chicago Press, 1946.

4. L.A. Pars, *Calculus of Variations*, Wiley, 1963.

5. R. Weinstock, *Calculus of Variations with Applications*, Dover, 1952.

6. I.M. Gelfand and S.V. Fomin, *Calculus of Variations*, Prentice-Hall, 1963.

7. T. Mura and T. Koya, *Variational Methods in Mechanics*, Oxford, 1992.

7.12 Worked Examples.

Example 7.N: Consider two given points (x_1, h_1) and (x_2, h_2), with $h_1 > h_2$, that are to be joined by a smooth wire. The wire is permited to have any shape, provided that it does not enter into the interior of the circular region $(x - x_0)^2 + (y - y_0)^2 \leq R^2$. A bead is released from rest from the point (x_1, h_1) and slides

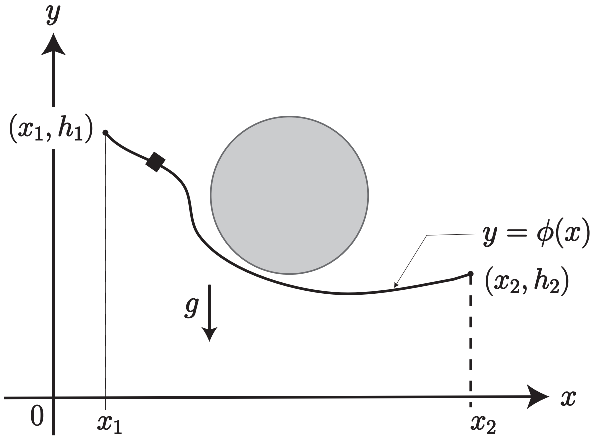


Figure 7.23: A curve $y = \phi(x)$ joining (x_1, h_1) to (x_2, h_2) which is disallowed from entering the forbidden region $(x - x_0)^2 + (\phi(x) - y_0)^2 < R^2$.

along the wire (without friction) due to gravity. For what shape of wire is the time of travel from (x_1, h_1) to (x_2, h_2) least?

Here the wire may not enter into the interior of the prescribed circular region . Therefore in considering different wires that connect (x_1, h_1) to (x_2, h_2), we may only consider those that lie entirely outside this region:

$$(x - x_0)^2 + (\phi(x) - y_0)^2 \geq R^2, \qquad x_1 \leq x \leq x_2. \tag{i}$$

The travel time of the bead is again given by (7.1) and the test functions must satisfy the same requirements as in the first example except that, in addition, they must be such satisfy the (inequality) constraint (i). Our task is to minimize $T\{\phi\}$ over the set A_1 subject to the constratint (i).

Example 7.N: Buckling: Consider a beam whose centerline occupies the interval $y = 0$, $0 < x < L$, in an undeformed configuration. A compressive force P is applied at $x = L$ and the beam adopts a buckled shape described by $y = \phi(x)$. Figure NNN shows the centerline of the beam in both the undeformed and deformed configurations. The beam is fixed by a pin at $x = 0$; the end $x = L$ is also pinned but is permitted to move along the x-axis. The prescribed geometric boundary conditions on the deflected shape of the beam are thus

$$\phi(0) = \phi(L) = 0.$$

By geometry, the curvature $\kappa(x)$ of a curve $y = \phi(x)$ is given by

$$\kappa(x) = \frac{\phi''(x)}{[1 + (\phi'(x))^2]^{3/2}}.$$

From elasticity theory we know that the bending energy per unit length of a beam is $(1/2)M\kappa$ and that the bending moment M is related to the curvature κ by $M = EI\kappa$ where EI is the bending stiffness of the beam. Thus the bending energy associated with a differential element of the beam is $(1/2)EI\kappa^2\, ds$ where ds

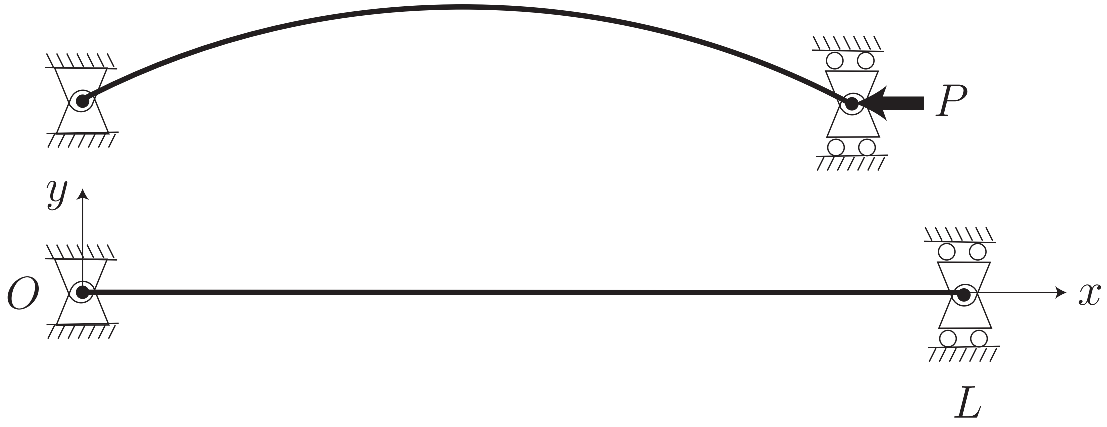


Figure 7.24: An elastic beam in undeformed (lower figure) and buckled (upper figure) configurations.

is arc length along the deformed beam. Thus the total bending energy in the beam is

$$\int_0^L \frac{1}{2} EI\,\kappa^2(x)\;ds$$

where the arc length s is related to the coordinate x by the geometric relation

$$ds \;=\; \sqrt{1+(\phi'(x))^2}\;dx.$$

Thus the total bending energy of the beam is

$$\int_0^L \frac{1}{2} EI\,\frac{(\phi'')^2}{[1+(\phi')^2]^{5/2}}\;dx.$$

Next we need to account for the potential energy associated with the compressive force P on the beam. Since the change in length of a differential element is $ds - dx$, the amount by which the right hand end of the beam moves leftwards is

$$-\Big(\int_0^L ds \;-\; \int_0^L dx\Big) \;=\; -\Big(\int_0^L \sqrt{1+(\phi')^2}\;dx \;-\; L\Big).$$

Thus the potential energy associated with the applied force P is

$$-\,P\left(\int_0^L \sqrt{1+(\phi')^2}\;dx \;-\; L\right).$$

Therefore the total potential energy of the system is

$$\Phi\{x,\phi,\phi',\phi''\} = \int_0^L \frac{1}{2} EI\,\frac{(\phi'')^2}{[1+(\phi')^2]^{5/2}}\;dx - \int_0^L P\left(\sqrt{1+(\phi')^2} \;-\; 1\right)dx.$$

The Euler equation, which for such a functional has the general form

$$\frac{d^2}{dx^2}\Big(f_{\phi''}\Big) \;-\; \frac{d}{dx}\Big(f_{\phi'}\Big) \;+\; f_\phi \;=\; 0,$$

simplifies in the present case since f does not depend explicitly on ϕ. The last term above therefore drops out and the resulting equation can be integrated once immediately. This eventually leads to the Euler equation

$$\frac{d}{dx}\left(\frac{\phi''}{[1+(\phi')^2]^{5/2}}\right) + \frac{\phi'}{2[1+(\phi')^2]^{1/2}}\left(\frac{P}{EI/2} \ + \ 5\left[\frac{\phi''}{[1+(\phi')^2]^{3/2}}\right]^2\right) = c$$

where c is a constant of integration, and the natural boundary conditions are

$$\phi''(0) = \phi''(L) = 0.$$

Example 7.N: Linearize BVP in buckling problem above. Also, approximate the energy and derive Euler equation associated with it.

Example 7.N: $u(x,t)$ where $0 \leq x \leq L, \ 0 \leq t \leq T$ Functional

$$F\{u\} = \int_0^T \int_0^L \Big(\frac{1}{2}u_t^2 \ - \frac{1}{2}u_x^2\Big) dxdt$$

Euler equation (Wave equation)

$$u_{tt} - u_{xx} = 0.$$

Example 7.N: Physical example? Functional

$$F\{u\} = \int_0^T \int_0^L \left(\frac{1}{2}u_t^2 \ - \Big(\frac{1}{2}u_x^2 + \frac{1}{2}m^2u^2\Big)\right) dxdt$$

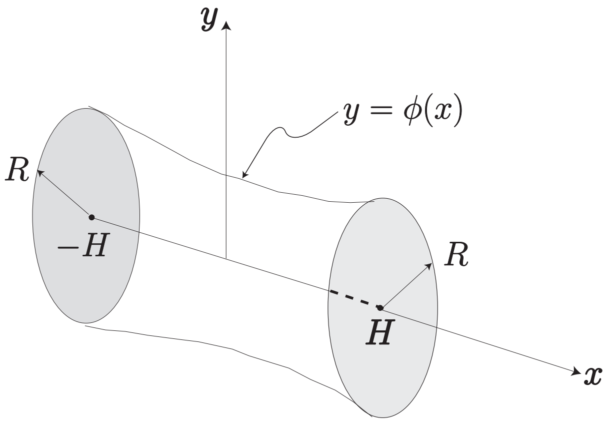
Euler equation (Klein-Gordon equation)

$$u_{tt} - u_{xx} + m^2u = 0.$$

Example 7.N: Null lagrangian

Example 7.2: Soap Film Problem. Consider two circular wires, each of radius R, that are placed coaxially, a distance H apart. The planes defined by the two circles are parallel to each other and perpendicular to their common axis. This arrangement of wires is dipped into a soapy bath and taken out. Determine the shape of the soap film that forms.

We shall assume that the soap film adopts the shape with minimum surface energy, which implies that we are to find the shape with minimum surface area. Suppose that the film spans across the two circular wires.



By symmetry, the surface must coincide with the surface of revolution of some curve $y = \phi(x), -H \le x \le H$. By geometry, the surface area of this film is

$$\text{Area}\{\phi\} = 2\pi \int \phi(x) ds = 2\pi \int_{-H}^{H} \phi(x)\sqrt{1 + (\phi')^2} dx,$$

where we have used the fact that $ds = \sqrt{1 + (\phi')^2} dx$, and this is to be minimized subject to the requirements $\phi(-H) = \phi(H) = R$ and

$$\phi(x) \ge 0 \qquad \text{for} \quad -H < x < H.$$

In order to determine the shape that minimizes the surface area we calculate its first variation δArea and set it equal to zero. This gives the Euler equation

$$\frac{d}{dx}\left\{\frac{\phi\phi'}{\sqrt{1+(\phi')^2}}\right\} - \sqrt{1+(\phi')^2} = 0$$

which we can write as

$$\frac{\phi\phi'}{\sqrt{1+(\phi')^2}} \frac{d}{dx}\left\{\frac{\phi\phi'}{\sqrt{1+(\phi')^2}}\right\} - \phi\phi' = 0$$

or

$$\frac{d}{dx}\left(\frac{\phi\phi'}{\sqrt{1+(\phi')^2}}\right)^2 - \frac{d}{dx}(\phi)^2 = 0.$$

This can be integrated to give

$$(\phi')^2 = \left(\frac{\phi}{c}\right)^2 - 1$$

where c is a constant. Integrating again and using the boundary conditions $\phi(H) = \phi(-H) = R$, leads to

$$\phi(x) = c\cosh\left(\frac{x}{c}\right) \tag{i}$$

where c is to be determined from the algebraic equation

$$\cosh\frac{H}{c} = \frac{R}{c}. \tag{ii}$$

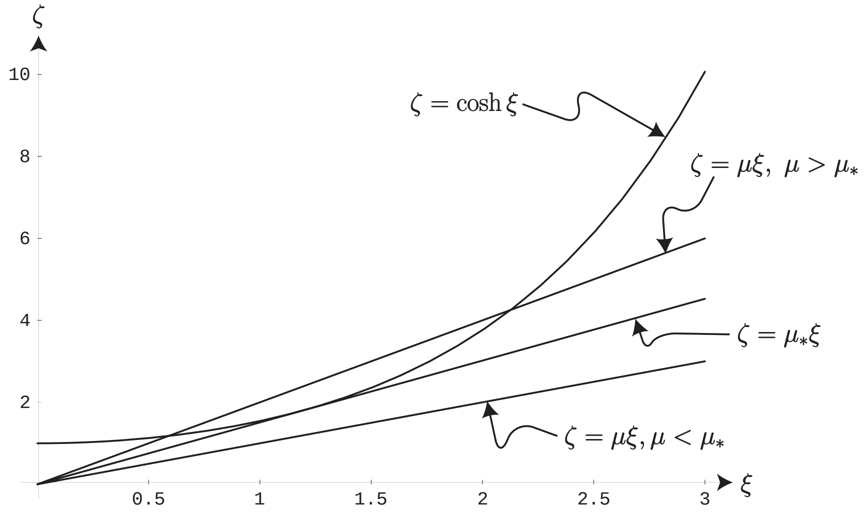


Figure 7.25: Intersection of the curve described by $\zeta = \cosh\xi$ with the straight line $\zeta = \mu\xi$.

Given H and R, if this equation can be solved for c, then the minimizing shape is given by (i) with this value (or values) of c.

To examine the solvability of (ii) set $\xi = H/c$ and $\mu = R/H$ and then this equation can be written as

$$\cosh\xi = \mu\xi.$$

As seen from Figure 7.25, the graph $\zeta = \cosh\xi$ intersects the straight line $\zeta = \mu\xi$ twice if $\mu > \mu_*$; once if $\mu = \mu_*$; and there is no intersection if $\mu < \mu_*$. Here $\mu_* \approx 1.50888$ is found by solving the pair of algebraic equations $\cosh\xi = \mu_*\xi,\ \sinh\xi = \mu_*$ where the latter equation reflects the tangency of the two curves at the contact point in this limiting case.

Thus in summary, if $R < \mu_* H$ there is no shape of the soap film that extremizes the surface area; if $R = \mu_* H$ there is a unique shape of the soap film given by (i) that extremizes the surface area; if $R > \mu_* H$ there are two shapes of the soap film given by (i) that extremize the surface area (and further analysis investigating the stability of these configurations is needed in order to determine the physically realized shape).

Remark: In order to understand what happens when $R < \mu_* H$ consider the following heuristic argument. There are three possible configurations of the soap film to consider: one, the film bridges across from one circular wire to the other but it does not form on the flat faces of the two circular wires themselves (which is the case analyzed above); two, the film forms on each circular wire but does not bridge the two wires; and three, the film does both of the above. We can immediately discard the third case since it involves more surface area than either of the first two cases. Consider the first possibility: the soap film spans across the two circular wires and, as an approximation, suppose that it forms a circular cylinder of radius R and length $2H$. In this case the area of the soap film is $2\pi R(2H)$. In the second case, the soap film covers only the two end regions formed by the circular wires and so the area of the soap film is $2 \times \pi R^2$. Since $4\pi RH < 2\pi R^2$ for $2H < R$, and $4\pi RH > 2\pi R^2$ for $2H > R$, this suggests that the soap film will span across the two circular wires if $R > 2H$, whereas the soap will not span across the two circular wires if $R < 2H$ (and would instead cover only the two circular ends).

Example 7.4: Minimum Drag Problem. Consider a space-craft whose shape is to be designed such that the drag on it is a minimum. The outer surface of the space-craft is composed of two segments as shown in Figure 7.26: the inner portion ($x = 0$ with $0 < y < h_1$ in the figure) is a flat circular disk shaped nose of radius h_1, and the outer portion is obtained by rigidly rotating the curve $y = \phi(x)$, $0 < x < 1$, about the x-axis. We are told that $\phi(0) = h_1, \phi(1) = h_2$ with the value of h_2 being given; the value of h_1 however is *not* prescribed and is to be chosen along with the function $\phi(x)$ such that the drag is minimized.

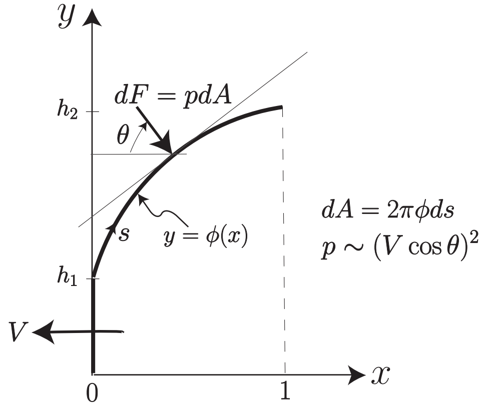


Figure 7.26: The shape of the space craft with minimum drag is generated by rotating the curve $y = \phi(x)$ about the x-axis. The space craft moves at a speed V in the $-x$-direction.

According to the most elementary model of drag (due to Newton), the pressure at some point on a surface is proportional to the square of the normal speed of that point. Thus if the space craft has speed V relative to the surrounding medium, the pressure on the body at some generic point is proportional to $(V\ \cos\theta)^2$ where θ is the angle shown in Figure 7.26; this acts on a differential area $dA = 2\pi y ds = 2\pi\phi ds$. The horizontal component of this force is therefore obtained by integrating $dF\cos\theta = (V\ \cos\theta)^2 \times (2\pi\phi ds) \times \cos\theta$ over the entire body. Thus the drag D is given, in suitable units, by

$$D = \pi h_1^2 + 2\pi \int_0^1 \frac{\phi(\phi')^3}{[1 + (\phi')^2]}\ dx,$$

where we have used the fact that $ds = dx\sqrt{1 + (\phi')^2}$ and $\cos\theta = \phi'/\sqrt{1 + (\phi')^2}$.

To optimize this we calculate the first variation of D, remembering that both the function ϕ and the parameter h_1 can be varied. Thus we are led to

$$\begin{aligned} \delta D &= 2\pi h_1 \delta h_1 \ + 2\pi \int_0^1 \frac{(\phi')^3}{[1 + (\phi')^2]} \delta\phi\ dx \ + \ 2\pi \int_0^1 \phi \left[\frac{3(\phi')^2}{1 + (\phi')^2} \ - \ \frac{2(\phi')^4}{[1 + (\phi')^2]^2} \right] \delta\phi'\ dx, \\ &= 2\pi h_1 \delta h_1 \ + 2\pi \int_0^1 \frac{(\phi')^3}{[1 + (\phi')^2]} \delta\phi\ dx \ + \ 2\pi \int_0^1 \left[\frac{\phi(\phi')^2(3 + (\phi')^2)}{[1 + (\phi')^2]^2} \right] \delta\phi'\ dx. \end{aligned}$$

Integrating the last term by parts and recalling that $\delta\phi(1) = 0$ and $\delta\phi(0) = \delta h_1$ (since the value of $\phi(1)$ is prescribed but the value of $\phi(0) = h_1$ is not) we are led to

$$\delta D = 2\pi h_1 \delta h_1 + 2\pi \int_0^1 \frac{(\phi')^3}{[1+(\phi')^2]}\delta\phi\, dx - 2\pi \int_0^1 \frac{d}{dx}\left[\frac{\phi(\phi')^2(3+(\phi')^2)}{[1+(\phi')^2]^2}\right]\delta\phi\, dx - \left.\frac{2\pi\phi(\phi')^2(3+(\phi')^2)}{[1+(\phi')^2]^2}\right|_{x=0}\delta h_1.$$

The arbitrariness of $\delta\phi$ and δh_1 now yield

$$\begin{array}{c} \dfrac{d}{dx}\left[\dfrac{\phi(\phi')^2(3+(\phi')^2)}{[1+(\phi')^2]^2}\right] - \dfrac{(\phi')^3}{[1+(\phi')^2]} = 0 \qquad \text{for} \qquad 0 < x < 1, \\ \left.\dfrac{(\phi')^2(3+(\phi')^2)}{[1+(\phi')^2]^2}\right|_{x=0} = 1. \end{array} \tag{i}$$

The differential equation in (i)$_1$ and the natural boundary condition (i)$_2$ can be readily reduced to

$$\begin{array}{c} \dfrac{d}{dx}\left(\dfrac{\phi(\phi')^3}{[1+(\phi')^2]^2}\right) = 0 \qquad \text{for} \qquad 0 < x < 1, \\ \phi'(0) = 1. \end{array} \tag{ii}$$

The differential equation (ii) tells us that

$$\frac{\phi(\phi')^3}{[1+(\phi')^2]^2} = c_1 \qquad \text{for} \qquad 0 < x < 1, \tag{iii}$$

where c_1 is a constant. Together with the given boundary conditions, we are therefore to solve the differential equation (iii) subject to the conditions

$$\phi(0) = h_1, \quad \phi(1) = h_2, \quad \phi'(0) = 1, \tag{iv}$$

in order to find the shape $\phi(x)$ and the parameter h_1.

Since $\phi(0) = h_1$ and $\phi'(0) = 1$ this shows that $c_1 = h_1/4$.

It is most convenient to write the solution of (iii) with $c_1 = h_1/4$ parametrically by setting $\phi' = \xi$. This leads to

$$\left.\begin{array}{rcl} \phi &=& \dfrac{h_1}{4}\Big(\xi^{-3} + 2\xi^{-1} + \xi\Big), \\ x &=& \dfrac{h_1}{4}\Big(\dfrac{3}{4}\xi^{-4} + \xi^{-2} + \log\xi\Big) + c_2, \end{array}\right\} \qquad 1 > \xi > \xi_2,$$

where c_2 is a constant of integration and ξ is the parameter. On physical grounds we expect that the slope ϕ' will decrease with increasing x and so we have supposed that ξ decreases as x increases; thus as x increases from 0 to 1 we have supposed that ξ decreases from ξ_1 to ξ_2 (where we know that $\xi_1 = \phi'(0) = 1$). Since $\xi = \phi' = 1$ when $x = 0$ the preceding equation gives $c_2 = -7h_1/16$. Thus

$$\left.\begin{array}{rcl} \phi &=& \dfrac{h_1}{4}\Big(\xi^{-3} + 2\xi^{-1} + \xi\Big), \\ x &=& \dfrac{h_1}{4}\Big(\dfrac{3}{4}\xi^{-4} + \xi^{-2} + \log\xi - \dfrac{7}{4}\Big), \end{array}\right\} \qquad 1 > \xi > \xi_2. \tag{v}$$

The boundary condition $\phi = h_1, \phi' = 1$ at $x = 0$ has already been satisfied. The boundary condition $\phi = h_2, x = 1$ at $\xi = \xi_2$ requires that

$$\left.\begin{aligned} h_2 &= \frac{h_1}{4}\Big(\xi_2^{-3} + 2\xi_2^{-1} + \xi_2\Big), \\ 1 &= \frac{h_1}{4}\Big(\frac{3}{4}\xi_2^{-4} + \xi_2^{-2} + \log\xi_2 - \frac{7}{4}\Big), \end{aligned}\right\} \tag{vi}$$

which are two equations for determining h_2 and ξ_2. Dividing the first of (vi) by the second yields a single equation for ξ_2:

$$\xi_2^5 - h_2\xi_2^4\log\xi_2 + \frac{7}{4}h_2\xi_2^4 + 2\xi_2^3 - h_2\xi_2^2 + \xi_2 - \frac{3}{4}h_2 = 0. \tag{vii}$$

If this can be solved for ξ_2, then either equation in (vi) gives the value of h_1 and (v) then provides a parametric description of the optimal shape. For example if we take $h_2 = 1$ then the root of (vii) is $\xi_2 \approx 0.521703$ and then $h_1 \approx 0.350943$.

Example 7.5: Consider the variational problem where we are asked to minimize the functional

$$F\{\phi\} = \int_0^1 f(\phi, \phi')dx$$

over some admissible set of functions A. Note that this is *a special case of the standard problem* where the function $f(x, \phi, \phi')$ is not explicitly dependent on x. In the present case f depends on x only through $\phi(x)$ and $\phi'(x)$.

The Euler equation is given, as usual, by

$$\frac{d}{dx}f_{\phi'} \ - \ f_\phi = 0 \qquad \text{for} \quad 0 < x < 1.$$

Multiplying this by ϕ' gives

$$\phi'\frac{d}{dx}f_{\phi'} \ - \ \phi' f_\phi = 0 \qquad \text{for} \quad 0 < x < 1,$$

which can be written equivalently as

$$\left[\frac{d}{dx}(\phi' f_{\phi'}) - \phi'' f_{\phi'}\right] \ - \ \left[\frac{d}{dx}f - \phi'' f_{\phi'}\right] = 0 \qquad \text{for} \quad 0 < x < 1.$$

Since this simplifies to

$$\frac{d}{dx}\Big[\phi' f_{\phi'} - f\Big] = 0 \qquad \text{for} \quad 0 < x < 1,$$

it follows that in this special case the Euler equation can be integrated once to have the simplified form

$$\phi' f_{\phi'} - f = \text{constant} \qquad \text{for} \quad 0 < x < 1.$$

Remark: We could have taken advantage of this in, for example, the preceding problem.

Example 7.6: Elastic bar. The following problem arises when examining the equilibrium state of a one-dimensional bar composed of a nonlinearly elastic material. An equilibrium state of the bar is characterized by a displacement field $u(x)$ and the material of which the bar is composed is characterized by a potential $\widehat{W}(u')$. It is convenient to denote the derivative of W by σ,

$$\widehat{\sigma}(u') = \widehat{W}'(u'),$$

so that then $\sigma(x) = \widehat{\sigma}(u'(x))$ represents the stress at the point x in the bar. The bar has unit cross-sectional area and occupies the interval $0 \leq x \leq L$ in a reference configuration. The end $x = 0$ of the bar is fixed, so that $u(0) = 0$, a prescribed force P is applied at the end $x = L$, and a distributed force per unit length $b(x)$ is applied along the length of the bar.

An admissible displacement field is required to be continuous on $[0, L]$, piecewise continuously differentiable on $[0, L]$ and to conform to the boundary condition $u(0) = 0$. The total potential energy associated with any admissible displacement field is

$$V\{u\} = \int_0^L \widehat{W}(u'(x))dx \ - \ \int_0^L b(x)u(x)dx \ - \ Pu(L),$$

which can be written in the conventional form

$$V\{u\} = \int_0^L f(x, u, u')dx \qquad \text{where} \quad f(x, u, u') = \widehat{W}(u') - bu - Pu'.$$

The actual displacement field minimizes the potential energy V over the admissible set, and so the three basic ingredients of the theory can now be derived as follows:

i. At any point x at which the displacement field is smooth, the Euler equation

$$\frac{d}{dx}\left(\frac{\partial f}{\partial u'}\right) - \frac{\partial f}{\partial u} = 0$$

takes the explicit form

$$\frac{d}{dx}\widehat{W}'(u') + b = 0,$$

which can be written in terms of stress as

$$\frac{d}{dx}\sigma + b = 0.$$

ii. The displacement field $u(x)$ satisfies the prescribed boundary condition $u = 0$ at $x = 0$. The natural boundary condition at the right hand end is given, according to equation (7.50) in Section 7.5.1, by

$$f_{u'} = 0 \qquad \text{at} \quad x = L,$$

which in the present case reduces to

$$\sigma(x) = \widehat{W}'(u'(x) = P \qquad \text{at} \quad x = L.$$

iii. Finally, suppose that u' has a jump discontinuity at some location $x = s$. Then the first Weirstrass-Erdmann corner condition (7.88) requires that $\partial f / \partial u'$ be continuous at $x = s$, i.e. that the stress $\sigma(x)$ must be continuous at $x = s$:

$$\sigma\big|_{x=s-} \ = \ \sigma\big|_{x=s+}. \tag{i}$$

The second Weirstrass-Erdmann corner condition (7.89) requires that $f - u'\partial f/\partial u'$ be continuous at $x = s$, i.e. that the quantity $W - u'\sigma$ must be continuous at $x = s$:

$$W - u'\sigma\Big|_{x=s-} \ = \ W - u'\sigma\Big|_{x=s+} \tag{ii}$$

Remark: The generalization of the quantity $W - u'\sigma$ to 3-dimensions in known as the Eshelby tensor.

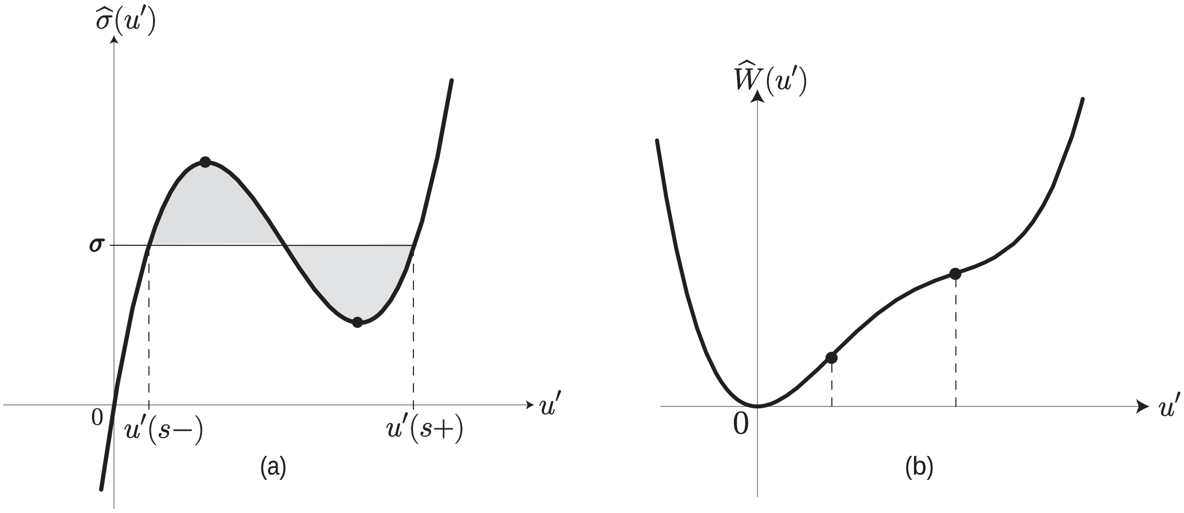


Figure 7.27: (a) A nonmonotonic (rising-falling-rising) stress response function $\widehat{\sigma}(u')$ and (b) the corresponding nonconvex energy $\widehat{W}(u')$.

In order to illustrate how a discontinuity in u' can arise in an elastic bar, observe first that according to the first Weirstrass-Erdmann condition (i), the stress σ on either side of $x = s$ has to be continuous. Thus if the function $\widehat{\sigma}(u')$ is monotonically increasing, then it follows that $\sigma = \widehat{\sigma}(u')$ has a unique solution u' corresponding to a given σ, and so u' must also be continuous at $x = s$. On the other hand if $\widehat{\sigma}(u')$ is a nonmonotonic function as, for example, shown in Figure 7.27(a), then more than one value of u' can correspond to the same value of σ, and so in such a case, even though $\sigma(x)$ is continuous at $x = s$ it is possible for u' to be discontinuous, i.e. for $u'(s-) \neq u'(s+)$, as shown in the figure. The energy function $\widehat{W}(u')$ sketched in Figure 7.27(b) corresponds to the stress function $\widehat{\sigma}(u')$ shown in Figure 7.27(a). In particular, the values of u' at which $\widehat{\sigma}$ has a local maximum and local minimum, correspond to inflection points of the energy function $\widehat{W}(u')$ since $\widehat{W}'' = 0$ when $\widehat{\sigma}' = 0$.

The second Weirstrass-Erdmann condition (ii) tells us that the stress σ at the discontinuity has to have a special value. To see this we write out (ii) explicitly as

$$\widehat{W}(u'(s+)) - u'(s+)\sigma \ = \ \widehat{W}(u'(s-) - u'(s-)\sigma$$

and then use $\widehat{\sigma}(u') = \widehat{W}'(u')$ to express it in the form

$$\int_{u'(s-)}^{u'(s+)} \widehat{\sigma}(v)dv = \sigma\big[u'(s+) - u'(s-)\big]. \tag{iii}$$

This implies that the value of σ must be such that the area under the stress response curve in Figure 7.27(a) from $u'(s-)$ to $u'(s+)$ must equal the area of the rectangle which has the same base and has height σ; or equivalently that the two shaded areas in Figure 7.27(a) must be equal.

Example 7.7: Non-smooth extremal. Find a curve that extremizes

$$F\{\phi\} = \int_0^1 f(x, \phi(x), \phi'(x))dx,$$

that begins from $(0, a)$, and ends at $(1, b)$ *after contacting a given curve* $y = g(x)$.

Remark: By identifying the curve $y = g(x)$ with the surface of a mirror and specializing the functional F to the travel time of light, one can thus derive the law of reflection for light.

Example 7.8: Inequality Constraint. Find a curve that extremizes

$$I(\phi) = \int_0^a (\phi'(x))^3 \, dx, \qquad \phi(0) = \phi(a) = 0,$$

that is prohibited from entering the interior of the circle

$$(x - a/2)^2 + y^2 = b^2.$$

Example 7.9: An example to caution against simple-minded discretization. (Due to John Ball). Let

$$F\{u\} = \int_0^1 (u^3(x) - x)^2 \, (u'(x))^2 dx$$

for all functions such that $u(0) = 0, u(1) = 1$. Clearly $F\{u\} \geq 0$. Moreover $F\{\bar{u}\} = 0$ for $\bar{u}(x) = x^{1/3}$. Therefore the minimizer of $F\{u\}$ is $\bar{u}(x) = x^{1/3}$.

Discretize the interval $[0, 1]$ into N segments, and take a piecewise linear test function that is linear on each segment. Calculate the functional F at this test function, next minimize it at fixed N, and finally take its limit as N tends to infinity. What do you get? (You will get an answer but not the correct one.)

To anticipate this difficulty in a different way, consider a 2 element discretization, and take the continuous test function

$$u_1(x) = \begin{cases} cx & \text{for} \quad 0 < x < h, \\ \bar{u}(x) & \text{for} \quad h < x < 1. \end{cases} \tag{i}$$

Calculate $F\{u\}$ for this function. Take limit as $h \to 0$ and observe that $F\{u\}$ does *not* go to zero (i.e. to $F\{\bar{u}\}$).

Example 7.10: Legendre necessary condition for a local minimum: Let

$$F\{\varepsilon\} = \int_0^1 f(x, \phi + \varepsilon\eta, \phi' + \varepsilon\eta')dx$$

for all functions $\eta(x)$ with $\eta(0) = \eta(1) = 0$. Show that

$$F''(0) = \int_0^1 \left\{ f_{\phi\phi}\eta^2 + 2f_{\phi\phi'}\eta\eta' + f_{\phi'\phi'}(\eta')^2 \right\} dx.$$

Suppose that $F''(0) \geq 0$ for all admissible functions η. Show that it is necessary that

$$f_{\phi'\phi'}(x, \phi(x), \phi'(x)) \geq 0 \qquad \text{for} \quad 0 \leq x \leq 1.$$

Example 7.11: Bending of a thin plate. Consider a thin rectangular plate of dimensions $a \times b$ that occupies the region $A = \{(x, y) \mid 0 < x < a,\ 0 < y < b\}$ of the x, y-plane. A distributed pressure loading $p(x, y)$ is applied on the planar face of the plate in the z-direction, and the resulting deflection of the plate in the z-direction is denoted by $w(x, y)$. The edges $x = 0$ and $y = 0$ of the plate are clamped, which implies the geometric restrictions that the plate cannot deflect nor rotate along these edges:

$$\left.\begin{aligned} w = 0, \quad \frac{\partial w}{\partial x} = 0 \qquad \text{on} \quad x = 0,\ 0 < y < b, \\ w = 0, \quad \frac{\partial w}{\partial y} = 0 \qquad \text{on} \quad y = 0,\ 0 < x < a; \end{aligned}\right\} \tag{i}$$

the edge $y = b$ is hinged, which means that its deflection must be zero but there is no geometric restriction on the slope:

$$w = 0, \qquad \text{on} \quad y = b,\ 0 < x < a; \tag{ii}$$

and finally the edge $x = a$ is free in the sense that the deflection and the slope are not geometrically restricted in any way.

The potential energy of the plate and loading associated with an admissible deflection field, i.e. a function $w(x, y)$ that obeys (i) and (i) is given by

$$\Phi\{w\} = \frac{D}{2} \int_A \left[\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} \right)^2 - 2(1 - \nu) \left(\frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2} - \left(\frac{\partial^2 w}{\partial x \partial y} \right)^2 \right) \right] dxdy - \int_A pw\ dxdy. \tag{iii}$$

where D and ν are constants. The actual deflection of the plate is given by the minimizer of Φ. We are asked to derive the Euler equation and the natural boundary conditions to be satisfied by minimizing w.

Answer: The Euler equation is

$$\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial x^4} = \frac{p}{D} \qquad 0 < x < a,\ 0 < y < b, \tag{iv}$$

and the natural boundary conditions are

$$\frac{\partial^2 w}{\partial y^2} + \nu\frac{\partial^2 w}{\partial x^2} = 0, \qquad \text{on} \quad y = b,\ 0 < x < a; \tag{v}$$

and

$$\frac{\partial^2 w}{\partial x^2}+\nu\frac{\partial^2 w}{\partial y^2}=0 \quad \text{and} \quad \frac{\partial^3 w}{\partial x^3}+2(1-\nu)\frac{\partial^3 w}{\partial x\partial y^2}=0, \qquad \text{on} \quad x=a,\ 0<y<b; \tag{vi}$$

Example 7.12: Consider the functional

$$F\{\phi\}=\int_0^1\Big[\big((\phi')^2-1\big)^2+\phi^2\Big]dx, \qquad \phi(0)=\phi(1)=0, \tag{i}$$

and determine a minimizing sequence $\phi_1,\phi_2,\phi_3,\ldots$ such that $F\{\phi_k\}$ approaches its infimum as $k\to\infty$. If the minimizing sequence itself converges to ϕ_* show that $F\{\phi_*\}$ is not the infimum of F.

Remark: Note that this functional is non-negative. If the functional takes the value zero, then, since its integrand is the sum of two non-negative terms, each of those terms must vanish individually. Thus we must have $\phi'(x)=\pm1$ and $\phi(x)=0$ on the interval $0<x<1$. These cannot be both satisfied by a regular function.

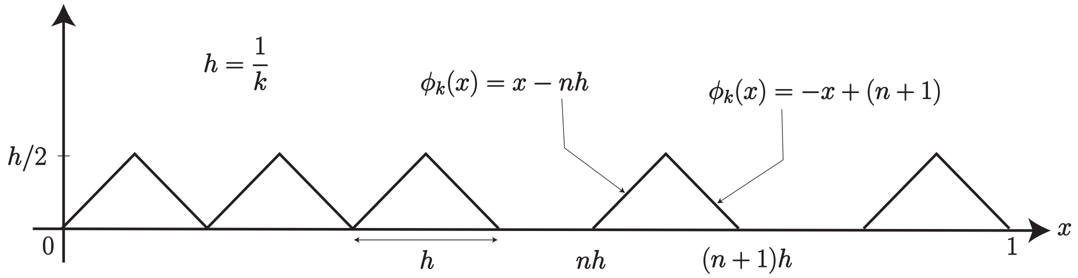


Figure 7.28: Sawtooth function $\phi_k(x)$ with k local maxima and linear segments of slope ±1.

Let $\phi_k(x)$ be the piecewise linear saw-tooth function with k-local maxima as shown in Figure 7.28; the slope of each linear segment is ±1. Note that the base $h=1/k$ and the height is $h/2$. Thus as k increases there are more and more teeth, each of which has a smaller base and smaller height. Observe that the first term in the integrand of (i) vanishes identically for any k; the second term, which equals the area under the square of ϕ_k approaches zero as $k\to\infty$. Thus the family of saw-teeth functions is a minimizing sequence of (i). However, note that since $\phi_k(x)\to0$ at each fixed x as $k\to\infty$ the limiting function to which this sequence converges, i.e. $\phi(x)=0$, is not a minimizer of (i).