ELECTROSTATIC LENS SYSTEMS

Second Edition

D W O HEDDLE

IOP

TSMC EX. 1016 Page 1 of 139

Electrostatic Lens Systems
Second Edition

Electrostatic Lens Systems
Second Edition

D W O Heddle
Emeritus Professor of Physics
Royal Holloway and Bedford New College
University of London

IoP
Institute of Physics Publishing
Bristol and Philadelphia

© IOP Publishing Ltd 2000

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher. Multiple copying is permitted in accordance with the terms of licences issued by the Copyright Licensing Agency under the terms of its agreement with the Committee of Vice-Chancellors and Principals.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN 0 7503 0697 1

Library of Congress Cataloging-in-Publication Data are available

Publisher: Nicki Dennis
Commissioning Editor: Jim Revill
Production Editor: Simon Laurenson
Production Control: Sarah Plenty
Cover Design: Victoria Le Billon
Marketing Executive: Colin Fenton

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK

US Office: Institute of Physics Publishing, The Public Ledger Building, Suite 1035, 150 South Independence Mall West, Philadelphia, PA 19106, USA

Typeset in TEX by Digital by Design, Cheltenham
Printed in the UK by Bookcraft, Midsomer Norton, Somerset

Contents

Preface **ix**

1 The optics of simple lenses **i**

1.1 Analogies between particle and photon optics 1
1.2 The cardinal points of a lens 4
1.2.1 The construction of an image 5
1.2.2 The general ray 8
1.2.3 The nodal points 9
1.3 Matrix representation of lens parameters 10

2 The motion of charged particles in an electrostatic field **13**

2.1 The equation of paraxial motion 13
2.2 Some general results 15
2.2.1 The nodal and principal points 17
2.2.2 Relativistic effects 18

3 The determination of the axial potential **20**

3.1 Analytic expressions 21
3.1.1 Approximations 23
3.1.2 The value of ω 25
3.2 The boundary element method 26
3.3 The finite element method 29
3.3.1 The basic equations 30
3.4 The finite difference method: relaxation 32
3.4.1 A two dimensional example 33
3.4.2 Cylindrical symmetry 38
3.4.3 Target conditions 40
3.4.4 Precision and accuracy 43
3.5 The nine point method applied to cylindrical geometry 46
3.5.1 The nine point formula for a general point 46
3.5.2 The nine point formula for an axial point 48

3.5.3 The convergence of the potential in the lens geometry 49
3.5.4 Comparisons of the five and nine point methods 50
3.6 Bessel function expansions 52
3.6.1 The variational principle 54
3.6.2 The problem of the non-linear potential in the gap 56
3.6.3 Calculation of the potential 57

4 The optics of simple lens systems 61

4.1 Aperture lenses 61
4.1.1 Three-aperture lenses 62
4.2 Cylinder lenses 64
4.3 The calculation of ray paths 65
4.3.1 Ray tracing in a known potential 65
4.3.2 Integration of the Picht equation 69
4.4 Real and asymptotic cardinal points 70
4.4.1 The determination of the cardinal points by ray tracing 71
4.5 Windows and pupils 71
4.5.1 The energy-add lens 73
4.6 Multi-element lenses 73
4.6.1 Zoom lenses 73
4.6.2 The afocal lens 75
4.6.3 The zoom afocal lens 77

5 Aberrations 80

5.1 Spherical aberration 80
5.1.1 The axial displacement 82
5.1.2 The disc of minimum confusion 83
5.2 Chromatic aberration 84
5.3 Off-axis aberrations 86
5.4 Interrelations of the spherical aberration coefficients 87
5.4.1 The principal surfaces 88
5.5 The determination of aberration coefficients 89
5.5.1 Direct ray tracing 89
5.5.2 Perturbation of the paraxial solution 92
5.6 The aberrations of retarding lenses 98
5.6.1 Spherical aberration 98
5.6.2 Fifth order effects 99
5.6.3 Chromatic aberration 102
5.6.4 Figures of merit 103

6 The LENSYS program 105

6.1 The programs on the disk 105
6.1.1 Loading and running the programs 106

TSMC EX. 1016 Page 7 of 139

6.2 The screen display 106
6.2.1 The menu system 106
6.3 The Set Options option 107
6.4 The Lens Data option 107
6.4.1 Error conditions 108
6.5 The Imaging option 109
6.5.1 Types of solution 109
6.5.2 Aperture lenses 111
6.5.3 Cylinder lenses 111
6.5.4 The Afocal and Varimag lenses 112
6.5.5 The special lenses 114
6.6 The Print/Save option 115
6.6.1 The Save to Files option 115
6.6.2 The printing options 116
6.6.3 Printing under Windows 117
6.7 The auxiliary programs 117
6.7.1 The **RELAX51A.EXE** and **RELAX51P.EXE** programs 117
6.7.2 The **RELAXCYL.EXE** program 118

Appendix: Technical aspects of LENSYS 120

A.1 The programming language 120
A.2 Initialization 120
A.2.1 The effect of processor type on the speed of the program 120
A.3 Calculation of potentials 121
A.3.1 Aperture lenses 121
A.3.2 Cylinder lenses 121
A.3.3 Details of the calculations 121
A.4 Ray tracing for focal data 124
A.4.1 Accuracy and consistency 125
A.5 Solving the imaging problem 125
A.5.1 Finding the image 125
A.5.2 Solving the lens 126
A.5.3 The afocal lens 126
A.5.4 The Varimag lens 127

References 128

Preface

Interest in electrostatic lenses for the control of ion and (especially) electron beams is of long standing and has grown considerably in the past few decades. In addition, recent innovations in the production of low energy positrons have opened a whole new field of research for which electrostatic lenses are required. This book makes no attempt to be a comprehensive text on electron and ion optics, but addresses the need for data on simple lenses consisting of two or three apertures or cylinders, together with some more complex cylinder lenses which have interesting properties. There are a number of quite detailed and extensive texts, of which two recent books, by Hawkes and Kasper [1] and by Szilagyi [2], contain modern treatments of the subject and are well worth consulting. The book by Harting and Read [3] which has objectives similar to the present volume has, unfortunately, been out of print for some years.

The book itself forms an *introductory* text on electrostatic optics and is accompanied by a disk containing a program for the IBM personal computer. This program, which is described in detail in chapter 6 and an appendix, will calculate the focal and aberration properties of a range of lenses and will also allow the study of the imaging behaviour between conjugate planes. Ray paths are traced in the paraxial approximation. A graphical display (Hercules, CGA, EGA or VGA) is required to run the program, and a maths coprocessor is very desirable. Only non-relativistic energies are considered and the electrode geometries are all very simple and easily manufactured.

The text bears some resemblance to courses I have given to first year postgraduate students over a number of years. The **LENSYS** program, on the other hand, is the result of a pressing need for lens data in a readily usable form for research purposes, and includes a number of lenses which have been found to be particularly useful for the control of electron and positron beams.

It is a pleasure to acknowledge the contribution made by two former research students, Dr Roger Cook and Dr Tony Renau, to the development of the Bessel function expansion method and its application to reveal hitherto unsuspected interrelationships between the aberration coefficients. I am particularly indebted to my colleague, Dr Susan Kay, who has worked with the **LENSYS** program from its

earliest manifestation as a number of separate routines. Her suggestions of material to include or omit, and her assistance in developing a reasonably 'user-friendly' interface, have been invaluable.

The principal changes in this new edition are in chapters 3 and 5 though chapter 6 and the appendix have also been substantially revised to reflect the changes in the **LENSYS** program itself. Chapter 3 now includes an account of the finite element method for the solution of the potential problem as well as the boundary element, finite difference and Bessel function expansion methods, which have all been expanded. There is a particular emphasis on the FDM, particularly in the nine point form, because it is the easiest of the methods for the reader to develop from scratch. The full equations are derived and, with material in the appendix, allow efficient, convergent programs to be written.

While **LENSYS** remains a DOS program, the rise in the Windows operating system and the associated great increases in computer memory since the first edition have been taken into account.

The **LENSYS** program has been expanded to include the calculation of the coefficients of chromatic aberration and the Petzval integral. The solution of the imaging problem has been extended and now, as well as the ability to find a focusing voltage appropriate to a specified object/image pair, the image position can be found for a specified set of potentials even if it lies outside the displayed screen. Data calculated during the ray tracing process may now be saved to files for further study and screen copies *in colour* may now be made via the Windows clipboard.

D W O Heddle
6 July 2000

1 The optics of simple lenses

1.1 Analogies between particle and photon optics

The actions of lenses and mirrors in controlling beams of light are matters of everyday experience and it may be helpful to look at some systems which have similar behaviour in photon and particle optics. The simplest, shown in figure 1.1, is the plane boundary separating two regions which differ in some property. In the case of photon optics the important property is the refractive index.

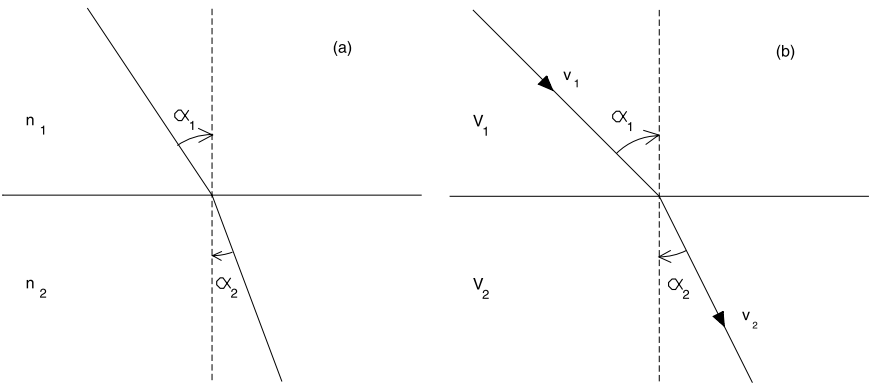

Figure 1.1 (a) Refraction of light at a plane boundary between two media having refractive indices n_1 and n_2; and (b) deviation of a beam of charged particles at a plane boundary separating regions having potentials V_1 and V_2.

The path of a ray of light incident non-normally onto the boundary is changed on crossing the boundary, the directions in the two regions being related by Snell's law. The analogue in particle optics is a boundary separating two regions at different electrostatic potentials. Here we consider a particle having a velocity v_1 in the first region and v_2 in the second. These velocities are related to the potentials, V_1

and V_2 in the two regions by

$$\tfrac{1}{2}mv_1^2 + eV_1 = \tfrac{1}{2}mv_2^2 + eV_2 = 0 \tag{1.1}$$

where e is the particle charge. This expression, which is fundamental to the simple analysis of electrostatic lenses, defines the zero of potential as that for which the particle is at rest. The potential changes abruptly at the boundary, but there is no change parallel to the boundary and so no force parallel to the boundary acts on the particle. The component of momentum parallel to the boundary is therefore unchanged and we can write

$$mv_1 \sin\alpha_1 = mv_2 \sin\alpha_2 \tag{1.2}$$

where α_1 and α_2 are the angles between the normal to the boundary and the path of the particle in the two regions. From equations (1.1) and (1.2) we have $(\sin\alpha_1)/(\sin\alpha_2) = (V_2/V_1)^{1/2}$ which is exactly Snell's law with $V^{1/2}$ playing the role of the refractive index.

That example was rather artificial, because abrupt changes of potential do not occur in free space, but can be only approximated by the use of closely spaced grids. Our second example is more realistic and demonstrates the focusing behaviour of a lens consequent on *curved* boundaries. Figure 1.2(a) depicts two regions, having refractive indices n_1 and n_2, separated by a spherical boundary of radius R. Rays of light incident parallel to the axis, at small distances, h_2 and h_1, from the right and left are refracted at the boundary in such a way that they cross the axis at points labelled F_1 and F_2 respectively. These are the first and second focal points of this simple lens and the distances measured from the boundary (and here we assume that $h \ll R$ so we can define the position of the boundary by the pole, O, where the boundary cuts the axis) are called the first and second focal lengths and denoted by f_1 and f_2 respectively. Remembering that h/R is very small, so the sines and tangents of the angles can be approximated by the angles themselves, we can write for the upper part of figure 1.2(a)

$$n_1\alpha_1 = n_2\alpha_2 \qquad \alpha_1 = h_1/R \qquad \alpha_1 - \alpha_2 = h_1/f_2$$

and for the lower part

$$n_1\alpha_4 = n_2\alpha_3 \qquad \alpha_3 = h_2/R \qquad \alpha_3 - \alpha_4 = h_2/f_1.$$

Eliminating the angles we find

$$\frac{1}{f_1} = \frac{(n_1 - n_2)}{n_1}\frac{1}{R} \qquad \frac{1}{f_2} = \frac{(n_2 - n_1)}{n_2}\frac{1}{R}. \tag{1.3}$$

Taking the ratio of these two expressions we see that the ratio of the two focal lengths is the negative of the ratio of the refractive indices. The sign is a

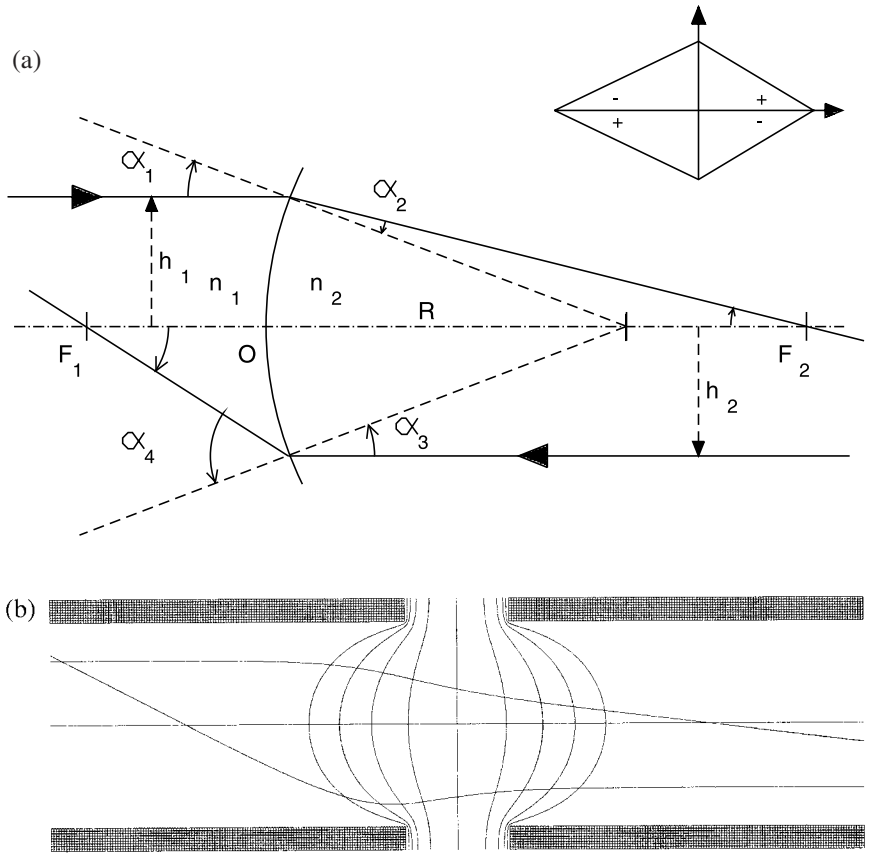

Figure 1.2 (a) Refraction of light at a spherical boundary between two media having refractive indices n_1 and n_2. The inset illustrates the sign convention. The angles α_1 and α_2 are positive as are the angles between the rays and the axis at the focal points. Since α_3 and α_4 are negative, the angle at F_1 is equal to $\alpha_3 - \alpha_4$. (b) Paths of charged particles in a two cylinder lens. The paths become asymptotic to straight lines only away from the central region of the lens. The values of the equipotential surfaces, shown in section, are discussed in the text.

consequence of defining the focal lengths as distances measured *from* the boundary and taking the pole to be the origin of Cartesian coordinates. The sign convention for angles is that angles measured from the axis to the ray are positive if the rotation is clockwise. These conventions lead to angles having the signs of their tangents, with the two distances involved measured *from* the right angle. One consequence of this choice is that the sign of the angle a ray makes with the axis is opposite to that of $\frac{dr}{dz}$ for that ray. For angles defined with respect to some other direction, such as a radius (as in figure 1.2(a)) or a second ray (as in figure 1.3), clockwise rotation

from the ray to the other direction defines the positive sense. In a number of the figures we have indicated the direction in which angles are measured according to these conventions.

Figure 1.2(b) illustrates an electrostatic analogue consisting of two coaxial cylinders separated by half their common diameter for clarity and held at potentials V_1 and V_2. No mathematical analysis should be needed to show that, deep inside each cylinder, the potentials are very close to those of the electrodes, but within a diameter or so of the centre of the system the potential changes quite rapidly and we show a number of equipotentials to illustrate this. These are symmetric about a flat, central, equipotential which has a value of $(V_2 + V_1)/2$ and those to the left have values $V_1 + (V_2 - V_1)/2^n$ where $n = 2, \ldots, 5$. Notice that the paths of the particles are curved and only away from the central region of the lens do they become asymptotic to straight lines. We shall have more to say about the curvature later, but for the moment we concentrate on the asymptotes. As in figure 1.2(a), rays entering the system parallel to the axis from the right and left are deviated towards the axis, crossing it at the first and second focal points, but as the deviation is not abrupt in this case there is no *physical* reference point from which to measure the focal lengths. Instead we consider the asymptotic paths and measure the focal lengths from the intersections of the asymptotes to the incident and emergent rays. These intersections define, for incident rays close to the axis, the *principal planes* of the lens. We shall see later that, for non-paraxial rays, the principal surfaces are actually curved; nonetheless the term 'principal plane' is universally employed. Notice that the actual paths are not perpendicular to the equipotential surfaces and approach the asymptotes from the outside on the high potential side of the lens. It would be possible to regard the equipotentials as boundaries separating regions having some mean potential and to use Snell's law to follow the refraction, boundary by boundary, through the lens, but this would be rather tedious and not very accurate. There are much better methods.

1.2 The cardinal points of a lens

The contrast between the ray paths in the photon and particle lenses of the previous section was very marked. In the case of a photon lens of many elements, such as is common in even quite simple cameras, the ray paths, while still made up of straight line segments, may experience many changes of direction within the lens and resemble somewhat the ray paths in the particle lens. The details of the paths are of little importance as long as the asymptotic paths are well defined and the appropriate planes of intersection known, because the imaging properties of both sorts of lens can be completely described (in the paraxial approximation, at least) by the positions of the focal points and the principal points, that is, the intersections of the principal planes with the axis. These are four of the six *cardinal points* of the lens and their use is illustrated in figure 1.3. The lens is specified in this figure by these four points, and their separations and the distances from an arbitrary

TSMC EX. 1016 Page 15 of 139

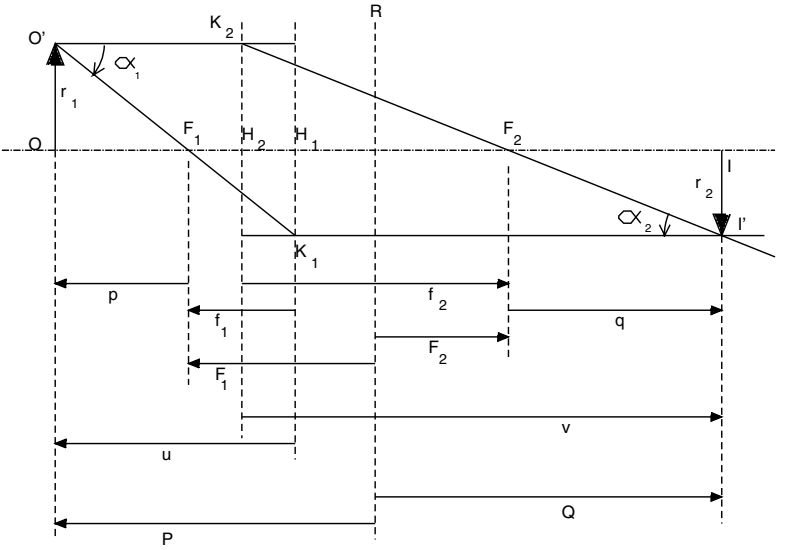

Figure 1.3 This diagram shows the focal and principal points of a lens and illustrates a construction, using asymptotes to the principal rays, for finding the position of the image of a given object. The angles α_1 and α_2 are measured from the upper ray to the lower and have opposite signs.

reference plane, R, are indicated. If there is a plane of (mechanical) symmetry in the lens, this plane is commonly used as reference and the distances from this plane to the focal points are called the 'mid-focal distances' and usually denoted by F_1 and F_2. The ambiguity of the same symbols being used for these distances and for the focal points themselves is not usually a problem.

Chapter 5 will contain a discussion of the aberrations of particle lenses, but until then we shall concentrate on the rather simpler situation of Gaussian imaging appropriate to paraxial rays, and in the next section shall develop some important interrelations. We have restricted the field of this book to lenses of axial symmetry, and we shall consider only rays which lie in planes containing the optic axis, that is, the meridional rays.

1.2.1 The construction of an image

We consider an object, OO′ of height r_1 at a distance p from the first focal point, u from the first principal point or P from the reference plane, and construct the image II′ by drawing two *principal rays* from O′. The first is drawn parallel to the axis to represent the initial asymptote and, from its intersection with the second principal plane at K_2, a line representing the asymptote to the emergent ray is drawn to pass through the second focal point, F_2. The second is drawn through the first focal point to meet the first principal plane at K_1 from where it continues parallel to the axis. The intersection of these two rays defines the image point, I′.

The image has a height r_2 and is a distance q from the second focal point, v from the second principal point or Q from the reference plane. With our Cartesian sign convention p, P, f_1, F_1, u and r_2 are all negative. The other distances are positive. The angles α_1 and α_2 have opposite signs because the rotations from one ray to the other are in different senses: we have chosen to mark α_1 as the positive angle.

In order to find relationships among the various distances, we consider a number of similar triangles. From triangles OF_1O' and $H_1F_1K_1$ we see that $r_1/p = -r_2/f_1$ and, from triangles IF_2I' and $H_2F_2K_2$, $r_2/q = -r_1/f_2$. Triangles $H_1F_1K_1$ and $H_2F_2K_2$ are not similar and allow us to represent the angles as $\alpha_1 = r_2/f_1$ and $\alpha_2 = -(-r_1/f_2)$. The (transverse) magnification of the lens, M, is given by the ratio r_2/r_1 and the angular magnification, M_α, by α_2/α_1 so we can write

$$M = -f_1/p = -q/f_2 \quad \text{and} \quad M_\alpha = (-r_1/r_2)(f_1/f_2)$$

so

$$MM_\alpha = -f_1/f_2 = n_1/n_2 = (V_1/V_2)^{1/2}. \tag{1.4}$$

This last relationship, between the magnification, angular magnification and the ratio of the potentials in the object and image regions, is known as the law of Helmholtz and Lagrange and is essentially a statement that the brightness of an image cannot exceed that of the object.

Using the two expressions for the magnification we obtain $pq = f_1 f_2$, which is Newton's relation, and if we substitute for u $(=p + f_1)$ and v $(=q + f_2)$ we obtain $f_1/u + f_2/v = 1$. This is not often used (except in an approximate form valid for thin lenses), but substitution for P $(=p + F_1)$ and Q $(=q + F_2)$ shows that a graph of Q against P is a rectangular hyperbola with its centre at F_1, F_2. This is frequently used to present data for lenses of two elements with the voltage ratio as a parameter. Lines of constant magnification can be shown on the same axes, leading to a very simple procedure for the choice of a (two element) lens to meet a specification in terms of a voltage ratio and magnification. In the same way that it is only the *ratio* of the potentials which governs the behaviour of a lens, so the actual size of the lens does not matter and it is only the ratios P/D, F_1/D etc where D is some characteristic dimension of the lens, usually the diameter, which are important. It is very common indeed to find focal lengths and other distances represented with D as the unit of length. In this book we shall normally write 'F_1', for example, instead of 'F_1/D' unless a specific point is to be made. Figure 1.4 illustrates the general form of a family of P–Q curves showing both branches of the hyperbolæ and also the real object-image region for a lens of two closely spaced coaxial cylinders.

With our sign convention, the transverse and angular magnifications of a real image are negative. It is, however, common usage to speak of magnifications as though they were positive and to ignore the sign when using terms such as 'larger', ' smaller', 'increase' etc. Except in critical situations we shall cite magnifications in this fashion.

TSMC EX. 1016 Page 17 of 139

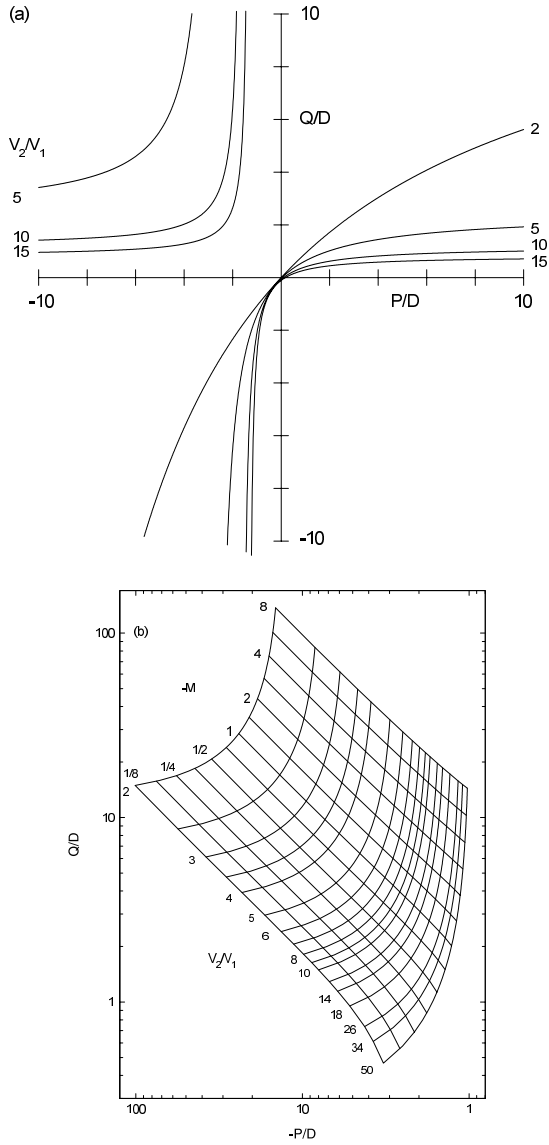

Figure 1.4 (a) P–Q diagram for a lens. The upper left region ($P < 0$, $Q > 0$) applies to the most common case of *real* objects and images. In the other regions one or both are virtual. Notice that, close to the origin, the P–Q relationship is virtually independent of the voltage ratio. (b) P–Q diagram for a two cylinder lens, with logarithmic scales to simplify the presentation of data over a wide range of distances and voltage ratios. Lines of constant magnification and of voltage ratio are shown.

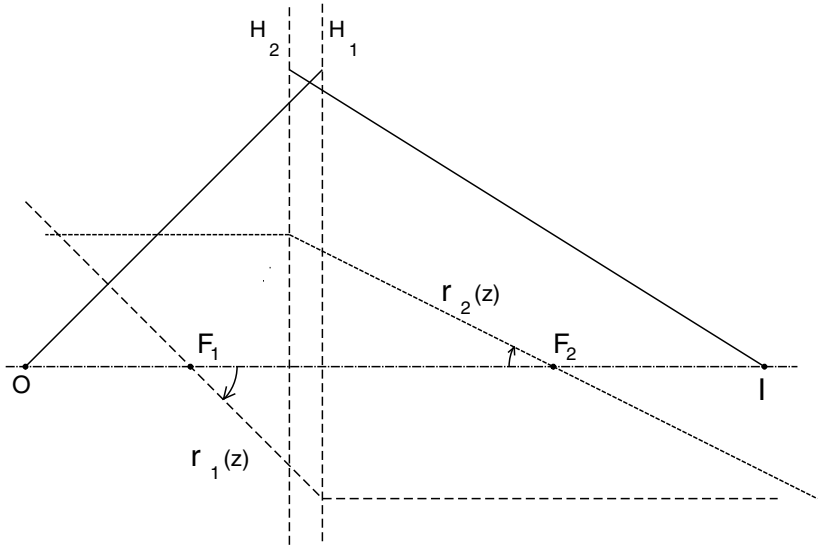

Figure 1.5 Asymptotes to two principal rays, $r_1(z)$ and $r_2(z)$, and to a general ray $r(z) = -[r_1(z) + (1/M)r_2(z)]$.

It is easy to see that an object in the first principal plane will be imaged in the second principal plane unchanged in size and the principal planes are sometimes referred to as 'planes of unit *positive* magnification'. Note that either the object or the image must be virtual in this case. We shall see later that this property of the principal planes has been exploited in particle optics.

1.2.2 The general ray

While the two principal rays shown in figure 1.3 offer a simple construction of the position and size of an image, a ray from an axial object point to the corresponding image point is often of greater interest. Provided that we restrict ourselves to paraxial, Gaussian, imaging, any linear combination of two rays which themselves satisfy the focusing condition will also satisfy that condition. Figure 1.5 illustrates a construction to demonstrate this property using the asymptotes to the various rays. Any independent pair of rays may be used as the basis, but it is convenient to choose rays which are parallel to the axis in one region. We define the first ray, $r_1(z)$, as that passing through the first focal point of the lens at an angle of $+45°$. This ray is therefore parallel to the axis in the image space and a distance f_1 from the axis. The second ray, $r_2(z)$, is parallel to the axis in the object space and a distance, $-f_1$, from the axis. It crosses the axis at the second focal point, with a slope f_1/f_2.

The general ray, $r(z)$, is then represented by $r(z) = \zeta r_1(z) + \xi r_2(z)$. The object is a distance p from the first focal point, so at the object we have $r(z) = \zeta(-p) + \xi(-f_1) = 0$, and we can therefore write $\xi = \zeta/M$ where M is the lens magnification between these conjugates. At the image we write

$r(z) = \zeta[f_1 + q(f_1/f_2)/M] = 0$ for all values of ζ because the term in the square brackets is zero. The single parameter, ζ, is sufficient to define a particular ray and we write

$$r(z) = \zeta\left(r_1(z) + \frac{1}{M}r_2(z)\right). \tag{1.5}$$

Figure 1.5 illustrates the construction of the asymptotes to one such general ray. We take $\zeta = -1$, and draw the asymptote from the object point, O, until it reaches the first principal plane, H_1. Then, from the second principal plane, H_2, at the same radial distance, draw the emergent asymptote to the image point, I. The slope of this asymptote is $(-1/M)f_1/f_2$ and, as M is known from the ratio of the first focal length and the object distance, the position of the image point does not have to be known beforehand. The values used for illustration in figure 1.5 are

$$f_1 = -4 \quad f_2 = 8 \quad p = -5 \quad q = 6.4$$

giving

$$M = -0.8.$$

1.2.3 The nodal points

The two remaining cardinal points are the *nodal points*, N_1 and N_2. These are points of unit positive *angular* magnification in the sense that a ray directed towards the first emerges as though from the second with the same slope. All rays from a point on the object must pass through the corresponding point on the image and we show a particular linear combination of the two principal rays in figure 1.6. We draw a ray for which the incident asymptote crosses the first principal plane at some distance h from the axis. The emergent asymptote appears to come from a point in the second principal plane at this same distance, h, from the axis. If these asymptotes are parallel we have

$$\frac{r_1 - h}{p + f_1} = \frac{r_2 - h}{q + f_2}$$

which leads to

$$h = r_1\left(\frac{f_1 + f_2}{f_2 - p}\right).$$

These asymptotes cross the axis at the nodal points which are clearly separated by the same distance as the principal points. We leave it as an exercise for the reader to show that these points are at distances f_2 and f_1 from the focal points F_1 and F_2 respectively. Notice that if the refractive indices (potentials) on the two sides of the lens are the same, i.e. $f_1 = -f_2$, then $h = 0$ and the asymptotes cross the axis at the principal points which, for this case, coincide with the nodal points.

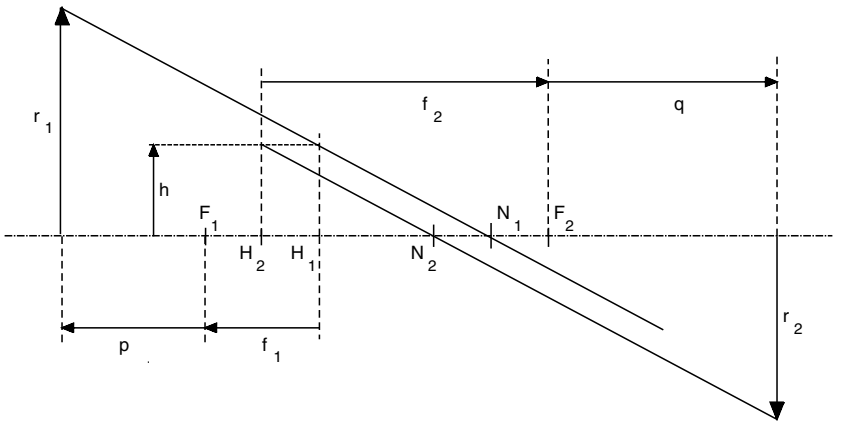

Figure 1.6 Diagram to illustrate the action at the nodal points of a lens.

1.3 Matrix representation of lens parameters

While the focal lengths and distances are the parameters most often presented, it is possible to express the action of a lens by a two by two matrix which acts on some initial ray asymptote for which the ray position, r, and slope, r', at some plane are expressed as a column vector, to produce a modified asymptote at a second plane.

$$\begin{pmatrix} r_2 \\ r_2' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} r_1 \\ r_1' \end{pmatrix}. \tag{1.6}$$

The planes between which the matrix acts determine the values of the matrix elements and some compromise is needed between simple expressions for the elements and a convenient choice of planes. The significance of the individual elements is clear from equation (1.6): a_{11} is the linear magnification between the planes and a_{22} the angular magnification. The other elements relate the radius at one plane and the slope at the other. The simplest matrix is one which transfers the ray from the first principal plane to the second. In this case the element $a_{11} = 1$ because the principal planes are planes of unit positive magnification, and $a_{22} = -\frac{f_1}{f_2}$ from equation (1.4). r_2 does not depend on r_1' so $a_{12} = 0$ and it is easy to show that $a_{21} = -\frac{1}{f_2}$ by considering an incident asymptote parallel to the axis.

The positions of the principal planes of an electrostatic lens vary with the potentials and it would be more convenient to work with a matrix operating between planes which did not move in this way. We now develop the matrix which transfers an incident asymptote at the reference plane to an emergent asymptote *at the same plane.* Despite the change in radial position which may occur, this is usually referred to as a *bending matrix.* We do this by operating on the incident column vector with three matrices. The first transfers the ray from the reference plane to the

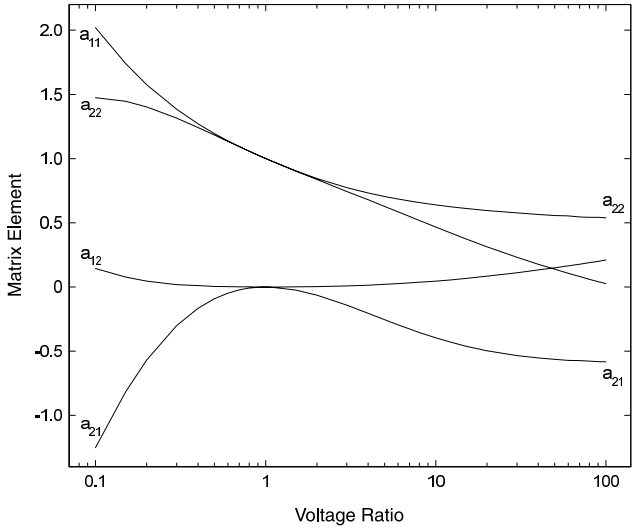

Figure 1.7 Matrix elements for a lens of two coaxial cylinders separated by 0.1 diameters. Note that all four matrix elements are continuous as the potential ratio passes through a value of one.

first principal plane, the second is the matrix acting between the principal planes and the third transfers the ray from the second principal plane to the reference plane.

$$\begin{pmatrix} r_2 \\ r_2' \end{pmatrix} = \begin{pmatrix} 1 & f_2 - F_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_2} & -\frac{f_1}{f_2} \end{pmatrix} \begin{pmatrix} 1 & F_1 - f_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_1' \end{pmatrix}$$

$$= \begin{pmatrix} \frac{F_2}{f_2} & \frac{F_1 F_2 - f_1 f_2}{f_2} \\ -\frac{1}{f_2} & -\frac{F_1}{f_2} \end{pmatrix} \begin{pmatrix} r_1 \\ r_1' \end{pmatrix}.$$

To use this matrix to study the formation of an image we first translate the ray from the object to the reference plane, a distance $-P$, then use the lens matrix to produce the emergent ray and finally transfer this ray a distance L. The product of these three matrices gives the *transfer matrix*

$$\begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{F_2}{f_2} & \frac{F_1 F_2 - f_1 f_2}{f_2} \\ -\frac{1}{f_2} & -\frac{F_1}{f_2} \end{pmatrix} \begin{pmatrix} 1 & -P \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{F_2 - L}{f_2} & \frac{(F_2 - L)(F_1 - P) - f_1 f_2}{f_2} \\ -\frac{1}{f_2} & \frac{P - F_1}{f_2} \end{pmatrix}.$$

The condition for imaging is that the *position* of the ray in the exit plane does not depend on the *slope* of the incident ray. This is equivalent to the element, a_{12}, of the transfer matrix being zero, which will be the case if $L = Q$. The element a_{11} is equal to the magnification of the lens between the conjugate planes, and a_{22} is the angular magnification.

The analysis of combinations of lenses is frequently simpler if the matrix formulation is used, because the only calculation required is a sequence of matrix multiplications. A further situation in which the matrix formulation has advantages is the analysis of very weak two element lenses. This is because the matrix elements vary smoothly as the potential ratio passes through one, in contrast to the focal lengths and distances which become infinite for this value of the ratio. Figure 1.7 shows the matrix elements for a lens of two coaxial cylinders separated by a gap of 0.1 diameters.

2 The motion of charged particles in an electrostatic field

2.1 The equation of paraxial motion

Our concern is with axisymmetric lenses and so we use cylindrical polar coordinates with the z-axis as the axis of symmetry of the lens system. Laplace's equation is

$$\frac{\partial^2 V}{\partial z^2} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) = 0 \tag{2.1}$$

where $V(r, z)$ is the potential. Because of the rotational symmetry an expansion of the potential in *even* powers of r can be made

$$V(r, z) = \sum_{n=0}^{\infty} A_n(z) r^{2n}.$$

The two terms of Laplace's equation are

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) = \sum_{n=0}^{\infty} 4n^2 A_n(z) r^{2n-2}$$

and

$$\frac{\partial^2 V}{\partial z^2} = \sum_{n=0}^{\infty} A_n''(z) r^{2n}$$

where primes ($'$) are used to denote differentiation with respect to z. The sum of the coefficients of each power of r must be zero so we have a recurrence relation

$$A_{n+1}(z) = -\frac{A_n''(z)}{4(n+1)^2}$$

giving the series expansion as

$$V(r, z) = A_0(z) - \frac{A_0''(z) r^2}{2^2} + \frac{A_0^{(4)}(z) r^4}{2^2 \cdot 4^2} + \cdots. \tag{2.2}$$

The axial potential, $V(z)$, is just $A_0(z)$ and we can write the axial and radial components of the electric field as

$$E_z = -\frac{\partial V}{\partial z} = -V' \qquad\qquad E_r = -\frac{\partial V}{\partial r} = \frac{r}{2}V''$$

where we omit the explicit reference to z and retain terms only to second order because we are considering only paraxial rays. When we come to consider the aberrations of particle lenses we shall have to make allowance for the higher order terms.

The axial velocity is so much greater than the radial in these circumstances that we can write $\frac{1}{2}m(\mathrm{d}z/\mathrm{d}t)^2 + eV = 0$ for the total energy of the particle. The equation of radial motion is

$$m\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = eE_r = \frac{er}{2}V''$$

writing e for the charge on the particle. The sign of the charge matters only insofar as the potential must always have the opposite sign and we shall soon see that the magnitude of the charge does not matter. To eliminate t from these expressions we write

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{\mathrm{d}z}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}z}{\mathrm{d}t}\frac{\mathrm{d}r}{\mathrm{d}z}\right)$$

giving

$$\left(\frac{-2eV}{m}\right)^{1/2}\frac{\mathrm{d}}{\mathrm{d}z}\left[\left(\frac{-2eV}{m}\right)^{1/2}\frac{\mathrm{d}r}{\mathrm{d}z}\right] = \frac{er}{2m}V''$$

which reduces to

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{V}\frac{\mathrm{d}r}{\mathrm{d}z} = -\frac{r}{4}\frac{V''}{V}. \tag{2.3}$$

This equation contains neither the charge of the particle nor its mass and it is therefore valid for electrons, positrons and ions of either sign. The only constraint is that the potential should have a sign opposite from that of the particle to ensure a positive total energy.

In order to integrate this equation of motion we require very precise information about the axial potential, because we need to determine the second derivative. An alternative approach is to change the independent variable in such a way as to remove this term. We introduce a *reduced radius*, R, defined by $R = rV^{1/4}$ where V must be read as $|V|$ for positive particles. Successive differentiation of this expression yields

$$\frac{\mathrm{d}R}{\mathrm{d}z} = V^{1/4}\frac{\mathrm{d}r}{\mathrm{d}z} + \frac{1}{4}V^{-3/4}V'r$$

$$\frac{\mathrm{d}^2 R}{\mathrm{d}z^2} = V^{1/4}\Bigg[\underbrace{\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{V}\frac{\mathrm{d}r}{\mathrm{d}z} + \frac{r}{4}\frac{V''}{V}}_{=0} - \frac{3r}{16}\left(\frac{V'}{V}\right)^2\Bigg]$$

where equation (2.3) shows that the grouped terms sum to zero, and we are left with the very simple equation of motion

$$\frac{\mathrm{d}^2 R}{\mathrm{d}z^2} = -\frac{3}{16} R \left(\frac{V'}{V}\right)^2 \tag{2.4}$$

which is known as the Picht equation [4].

2.2 Some general results

The independent variable in the Picht equation is the ratio of the axial potential gradient to the potential itself and it will be convenient to define a new variable, $T(z) = (V'(z)/V(z))$, and to note that, because this appears *squared* the sign does not matter.

Figure 2.1(a) shows the variation of the axial potential, the potential gradient and the parameter $T(z)$ for a lens having a gap equal to the radius with a potential ratio of 1 : 16. Figure 2.1(b) shows the lens geometry and equipotentials at values of $\frac{1}{32}$, $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, $\frac{7}{8}$, $\frac{15}{16}$ and $\frac{31}{32}$ of the overall potential difference. The potential distribution was calculated by the nine point finite difference method of section 3.5 and the two principal rays were traced in this lens by the routines of the **LENSYS2** program.

It is not difficult to show that all electrostatic lenses having uniform potential regions to each side are converging and to obtain an approximate expression for their focal lengths. We consider a ray incident parallel to the axis at a reduced radius R_1 and make the assumption that this reduced radius does not change in passing through the lens. Naturally, r, the true radius, will change or there would be no lens action, but the change in V will act in the opposite sense. A formal integration of the Picht equation gives

$$\int_{-\infty}^{\infty} R''\,\mathrm{d}z = R_2' - R_1' = -\frac{3}{16} R_1 \int_{-\infty}^{\infty} T^2\,\mathrm{d}z$$

though in practice the limits of integration can be very much narrower as a consequence of the sharply peaked nature of T^2. $R_1' = 0$ because the incident ray is parallel to the axis and so R_2' has the opposite sign to R_1. Writing

$$R_2' = r_2' V_2^{1/4} + \frac{r_2}{4} V_2^{-3/4} V_2'$$

and, noting that V_2' will be zero away from the lens proper and that $V_2^{1/4}$ is intrinsically positive, we see that r_2' has the same sign as R_2' and so for rays incident above the axis the emergent ray moves towards the axis giving a *convergent* lens action. Lenses for which the object and image positions lie in regions of uniform potential are known as *immersion lenses.*

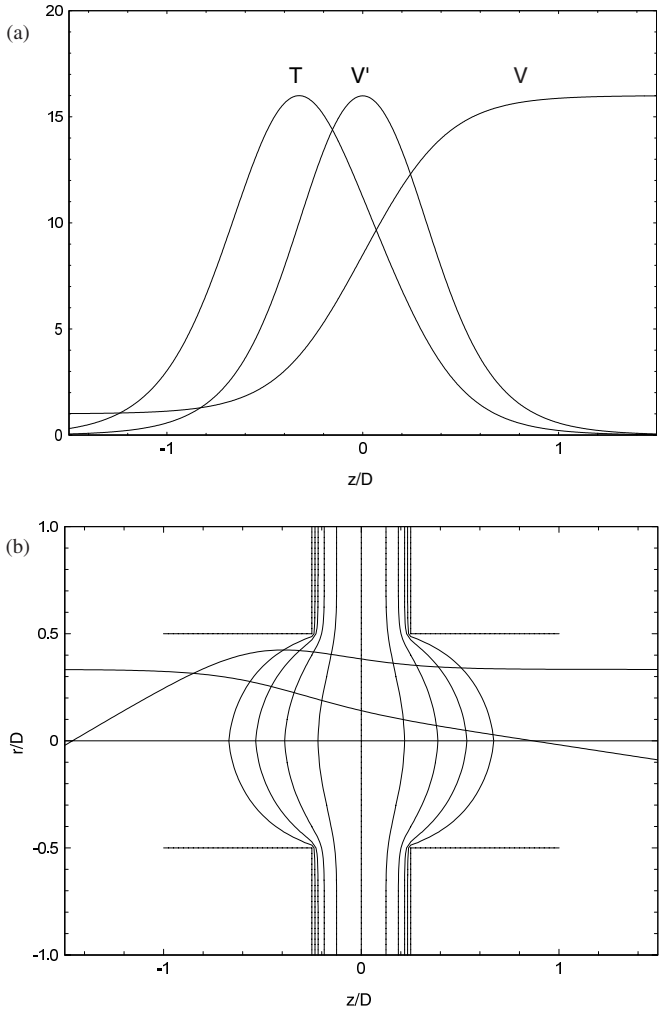

Figure 2.1 (a) The variation of $V(z)$, $V'(z)$ and their ratio, $T(z)$ with axial position z/D in a two cylinder lens having a gap to diameter ratio of 0.5. Note that the maximum value of T occurs on the low potential side of the lens. The ordinate scale applies to the axial potential. (b) Ray paths for a voltage ratio of 1 : 16 in this lens showing that most of the deviation of the particle paths occurs in the region where $|T|$ is large.

We saw in section 1.2.1 that the second focal length of a lens can be written as $-r_1/\alpha_2$. We identify α_2 with r_2' and write

$$\frac{1}{f_2} = -\frac{r_2'}{r_1} = -\frac{R_2'}{R_1}\left(\frac{V_1}{V_2}\right)^{1/4} = \frac{3}{16}\left(\frac{V_1}{V_2}\right)^{1/4}\int_{-\infty}^{\infty}\left(\frac{V'}{V}\right)^2 \mathrm{d}z. \qquad (2.5)$$

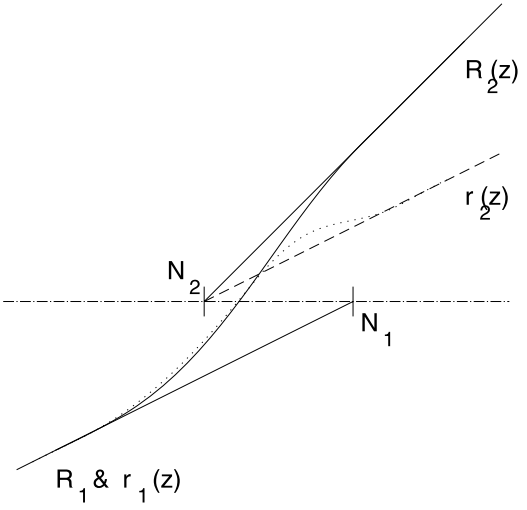

Figure 2.2 Paths of the ray $r(z)$ (· · · · · ·) and the reduced ray $R(z)$ (——) in the vicinity of the nodal points. The reduced ray is bounded by the asymptotes and the curvature of its path is always towards the axis. The slopes of the asymptotes to the ray, $r(z)$, are equal, while those of the reduced ray depend on the potential ratio. The radial scales have been chosen to make the incident asymptotes coincide.

If we were to trace a ray incident parallel to the axis, but from the other side, we would obtain

$$\frac{1}{f_1} = -\frac{r_1'}{r_2} = -\frac{R_1'}{R_2}\left(\frac{V_2}{V_1}\right)^{1/4} = \frac{3}{16}\left(\frac{V_2}{V_1}\right)^{1/4}\int_{\infty}^{-\infty}\left(\frac{V'}{V}\right)^2 \mathrm{d}z \tag{2.6}$$

and would find, for the ratio of the focal lengths,

$$\left(\frac{f_1}{f_2}\right) = -\left(\frac{V_1}{V_2}\right)^{1/2} = -\frac{n_1}{n_2}.$$

The ratio is negative because the integrals in equations (2.5) and (2.6) have opposite signs and the result is what we expect from our previous general discussion.

An alternative expression is sometimes quoted

$$\frac{1}{f_2} = \frac{1}{8V_2^{1/2}}\int_{-\infty}^{\infty}\frac{V'^2}{V^{3/2}}\,\mathrm{d}z$$

but the derivation of this expression is valid only if V'' does not change sign and this is never true unless the potential is constrained by grids.

2.2.1 The nodal and principal points

We consider an incident ray having its asymptote directed towards the first nodal point as illustrated in figure 2.2. By a suitable choice of vertical scale the same

line can be used to represent the asymptotes of both the ray itself and the reduced ray. The Picht equation shows that $R'' \propto -R$ and so the path of the reduced ray is always curved *towards* the axis. The reduced ray therefore crosses the axis to the left of the first nodal point; the curvature then reverses and the ray approaches its emergent asymptote from below. This asymptote passes through the second nodal point which must lie to the left of the point at which the ray crossed the axis. Note that both the 'real' ray, shown by the dotted line, and the reduced ray cross the axis at the same point. While the slope of the real asymptotes are equal, those of the reduced asymptotes are in the ratio of the one fourth power of the potential ratio. As figure 2.2 shows, the actual ray path is not bounded by the asymptotes. In chapter 1 we saw that the separation of the nodal points is equal to that of the principal points and our result, above, shows that the principal points of an electrostatic immersion lens are crossed, in the sense that H_2 is further to the low potential side of the lens than H_1. This is just the way in which the principal planes were represented in figure 1.3.

2.2.2 Relativistic effects

The independence of the paths of charged particles in electrostatic fields from the charge to mass ratio is only true for velocities sufficiently small that the mass of the particle remains essentially equal to the rest mass. Certainly the lenses discussed in chapter 6 and illustrated in the program would not be suitable for use with high potentials, but it is perhaps worth noting here the ways in which relativistic effects can be taken into account. Writing the momentum of a particle as

$$p = m_0 v(1 - \beta^2)^{1/2} \tag{2.7}$$

and the kinetic energy as

$$E = m_0 c^2((1 - \beta^2)^{-1/2} - 1) \tag{2.8}$$

the velocity can be expressed in terms of the momentum (using equation (2.7)) and substituted into equation (2.8) to give

$$m_0 c^2 + E = c\sqrt{m_0^2 c^2 + p^2}$$

which can be rearranged to give an explicit expression for the momentum in terms of the kinetic energy. Recalling that this is just equal to the negative of the charge multiplied by the potential (the zero being taken as usual for the particle to be at rest), we finally express the momentum in terms of the accelerating potential, V.

$$\begin{aligned} p^2 &= 2m_0 E\left(1 + \frac{E}{2m_0 c^2}\right) \\ &= -2m_0 eV\left(1 - \frac{eV}{2m_0 c^2}\right) \\ &= -2m_0 eV_{rel} \end{aligned}$$

where

$$V_{rel} = V\left(1 - \frac{eV}{2m_0c^2}\right) = V(1+\epsilon)$$
$$= V(1 + 0.9785 \times 10^{-6}V) \qquad \text{for electrons}$$

is known as the relativistically corrected acceleration potential. Note that the product eV is always negative, and therefore ϵ is a positive quantity.

The equations of motion can be analysed in terms of this corrected potential and the Picht equation retains much the same form. The reduced radius, R, is now defined by $R = rV_{rel}^{1/4} = rV^{1/4}(1+\epsilon)^{1/4}$ and the equation itself becomes

$$R'' + \frac{3}{16}R\left(\frac{V'}{V}\right)^2 \times K(\epsilon) = 0$$

where

$$K(\epsilon) = \frac{1 + \frac{4}{3}\epsilon + \frac{4}{3}\epsilon^2}{(1+\epsilon)^2}.$$

3 The determination of the axial potential

In order to trace the path of a particle through a lens, we have to integrate the equation of motion and it is convenient and simple to work in terms of the Picht equation. We need to know the function $T(z)$ to find the focal properties and we shall see later that we also need to know the axial derivative, $T'(z)$, if we wish to obtain information on the aberrations. The first stage in the calculation is the determination of the axial potential, from which, with its derivative, we construct $T(z)$. We shall consider ways of generating the potential distribution for lenses with more than two electrodes later, but in the first instance we concentrate our attention on lenses having only two electrodes. If the electrode potentials are V_1 and V_2, we can express the axial potential in one of the forms

$$\begin{aligned} V(z) &= \tfrac{1}{2}(V_1 + V_2) + \tfrac{1}{2}(V_2 - V_1)\Phi(z) \\ &= V_1 + (V_2 - V_1)\Psi(z) \end{aligned}$$

where $\Phi(z)$ and $\Psi(z)$ are functions which range between ∓ 1 and between 0 and 1 respectively as $z \to \mp\infty$.

There are various methods of calculating these functions. There is an exact solution for the case of two coaxial cylinders at different potentials, but with no gap between them. This is an unrealistic situation, but it has been used as an aid to the normalization of other expressions.

The general form can be approximated by a number of analytic expressions chosen more for their mathematical tractability than their connection with the basic physics of the problem. Such an approach has a long history, but with the amount of computing power readily available nowadays, there is little point in using this sort of method though we give a brief account in the next section. Results of much greater accuracy can be obtained by approximations based on the laws of electrostatics. In later sections we consider three such approaches: the direct application of Coulomb's law, numerical solutions of Laplace's equation on discrete networks of points and the analytical solution of Laplace's equation subject to boundary conditions which closely approximate the true situation.

TSMC EX. 1016 Page 31 of 139

3.1 Analytic expressions

An exact analytic solution of Laplace's equation can be found for the axisymmetric case, where the potential, $V(z, r)$, can be written as a constant term plus the product of separate functions, $Z(z)$ and $R(r)$ of the axial and radial coordinates. If we differentiate this product and substitute into Laplace's equation we obtain an equation in which the terms in z and in r may each be set equal to some parameter which we write as $-k^2$.

$$\frac{1}{R}\left(\frac{\mathrm{d}^2R}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}R}{\mathrm{d}r}\right) = -\frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} = -k^2.$$

The differential equations for the two functions are then

$$\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} - k^2Z = 0 \qquad \frac{\mathrm{d}^2R}{\mathrm{d}(kr)^2} + \frac{1}{kr}\frac{\mathrm{d}R}{\mathrm{d}(kr)} + R = 0.$$

The solution for $Z(z)$ is straightforward:

$$Z(z) = A_k\mathrm{e}^{kz} + B_k\mathrm{e}^{-kz}$$

while that for $R(r)$ requires some consideration of the context. The equation is Bessel's equation of zero order for which the formal solution would be

$$R(kr) = C_1J_0(kr) + C_2N_0(kr)$$

where J_0 and N_0 are Bessel functions of zero order and of the first and second kind respectively. The latter become infinite for zero argument and cannot therefore give a physically real description of the potential on the axis, so $C_2 = 0$ and the overall solution for the potential becomes

$$V(z, r) = (A_k\mathrm{e}^{kz} + B_k\mathrm{e}^{-kz})J_0(kr) + V_0 \tag{3.1}$$

where we have incorporated the coefficient C_1 into A_k and B_k and V_0 is a constant potential determined by the boundary conditions. Notice that the zero order Bessel function of the first kind can be expressed as

$$J_0(r) = \sum_{i=0}^{\infty}\frac{(-1)^i(r/2)^{2i}}{(i!)^2} = 1 - \frac{r^2}{4} + \frac{r^4}{64} - \frac{r^6}{2304} + \cdots$$

which is exactly the form given in equation (2.2) by the Taylor expansion. The general solution for a particular lens geometry will consist of a sum of such terms with, perhaps, a constant representing the potential of an electrode.

If k is a real number then, in equation (3.1), A_k must be set to zero for positive values of kz and B_k must be zero for negative values. In section 3.6 we shall use these distinct forms, but for a general case we have to assume that $k = \mathrm{j}\kappa$ and is an imaginary number. Using the modified Bessel function, $I_n(x) = \mathrm{j}^{-n} J_n(\mathrm{j}x)$, and expanding the exponentials of equation (3.1) we write

$$V(z, r) = \big(G_\kappa \cos(\kappa z) + F_\kappa \sin(\kappa z)\big) I_0(\kappa r) + V_0.$$

This is a solution of Laplace's equation for an axially symmetric field and for a specific value of κ. To find the general solution we must integrate over all possible values of κ.

$$V(z, r) = \int_{-\infty}^{\infty} \big(G_\kappa \cos(\kappa z) + F_\kappa \sin(\kappa z)\big) I_0(\kappa r)\, \mathrm{d}\kappa + V_0.$$

We consider the case of two coaxial cylinders of radius R separated by an infinitesimal gap and take the following values of the potentials to define the boundary conditions:

$$\begin{array}{ll} \text{for } z < 0 & V(z, R) = V_1 \\ \text{for } z = 0 & V(0, r) = (V_1 + V_2)/2 \\ \text{for } z > 0 & V(z, R) = V_2. \end{array}$$

The second of these conditions both determines the value of V_0 and requires that G_κ should be zero to remove the cosine term.

$$V(z, r) = \int_{-\infty}^{\infty} F_\kappa \sin(\kappa z) I_0(\kappa r)\, \mathrm{d}\kappa + \frac{V_1 + V_2}{2}. \tag{3.2}$$

Application of the first and third boundary conditions allows us to show, after an application of Fourier's integral theorem, that the coefficient F_k can be expressed as

$$\frac{(V_2 - V_1)}{2\pi\kappa I_0(\kappa R)}$$

and if we substitute this into equation (3.2), rearrange the terms and note that the integrand is an even function, we find that the potential can be written as

$$V(z, r) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{2} \frac{2}{\pi} \int_0^{\infty} \frac{\sin(\kappa z)}{\kappa} \frac{I_0(\kappa r)}{I_0(\kappa R)}\, \mathrm{d}\kappa \tag{3.3}$$

which reduces to the correct boundary values for $r = R$ as

$$\int_0^{\infty} \frac{\sin(\kappa z)}{\kappa}\, \mathrm{d}\kappa = -\frac{\pi}{2},\ 0,\ \frac{\pi}{2} \quad \text{for } z < 0,\ 0,\ > 0.$$

This expression is not amenable to evaluation except by numerical methods and there is only one reason for continuing the analysis.

3.1.1 Approximations

It was pointed out by Gray in 1939 [5] that the axial potential function for the two cylinder zero gap lens can be represented approximately by the function

$$\Phi(z) = \tanh \frac{\omega z}{R} \tag{3.4}$$

which has the correct asymptotic behaviour. The precise shape is governed by the parameter ω. If we differentiate this expression we have

$$\frac{\mathrm{d}\Phi}{\mathrm{d}z} = \frac{\omega}{R} \operatorname{sech}^2 \frac{\omega z}{R} = \frac{\omega}{R} \qquad \text{at} \quad z = 0.$$

From equation (3.3) we can extract the true shape of the axial potential as

$$\Phi(z) = \frac{2}{\pi R} \int_0^\infty \frac{\sin(\kappa z)}{\kappa} \frac{\mathrm{d}\kappa}{I_0(\kappa)} \tag{3.5}$$

and we can equate the differential of this at $z = 0$ to ω/R to find

$$\omega = \frac{2}{\pi} \int_0^\infty \frac{\mathrm{d}\kappa}{I_0(\kappa)}. \tag{3.6}$$

Cylinders with no gap between them are somewhat impracticable, but this approximate form for $\Phi(z)$ does give a reasonable guide to the behaviour of the lens with a small gap. Writing γ for the potential ratio V_1/V_2, the parameter, $T(z)$, of the Picht equation can be expressed as

$$T = \frac{2\omega(1-\gamma)\exp 2\omega z}{(1+\exp 2\omega z)(\gamma + \exp 2\omega z)} \tag{3.7}$$

which has a maximum value,

$$T_m = -2\omega\left(\frac{\gamma^{1/2} - 1}{\gamma^{1/2} + 1}\right)$$

at $z_m = (1/4\omega)\ln\gamma$. Analysis using this approximate potential function usually continues by replacing equation (3.7) by one of two expressions having a similar shape:

$$T = \frac{T_m}{1 + ((z - z_m)/a)^2} \tag{3.8}$$

or

$$T = T_m \operatorname{sech}((z - z_m)/a) \tag{3.9}$$

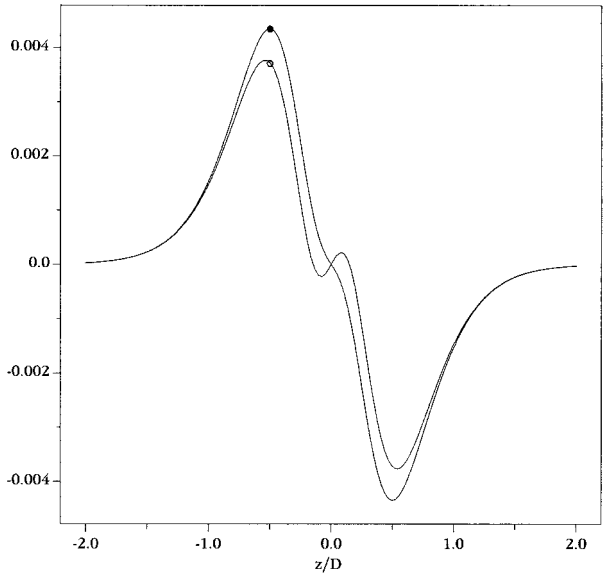

Figure 3.1 The differences between the potential functions, Φ, of equations (3.4) (●) and (3.10) (○) and the true axial potential for a two cylinder lens having a gap of $D/10$.

where a is found by equating the area under the curves of equations (3.7), (3.8) and (3.9). The focal lengths and distances can then be expressed as rather complicated functions of γ, a, ω and z_m.

For the more practical case of cylinders separated by a small, but not infinitesimal, gap, g, the function

$$\Phi = \frac{1}{\omega' g/R} \ln\left(\frac{\cosh(\omega z/R + \omega' g/2R)}{\cosh(\omega z/R - \omega' g/2R)}\right) \tag{3.10}$$

has often been used. If $\omega' = \omega$ this reduces to $\tanh \omega z/R$ as $g \to 0$. The expression for $T(z)$ is more complex and the approximation becomes progressively poorer as g increases. Some improvement may be made by modifying the value of ω' to take account of the smaller value of the slope of the potential at $z = 0$, and a value of 1.67 has been used, but this is not really a good model for a wide gap lens. It was in any case common practice to use equations (3.8) or (3.9) for actual calculations. Figure 3.1 shows the differences between the values of Φ given by these two approximations and the true axial potential for a two cylinder lens with a gap of $D/10$.

Some of the other approximate forms which have been suggested are:

$$\Psi = 1 - \frac{1}{\pi}\,\mathrm{arccot}\left(\frac{z}{a}\right) \qquad \Psi = \left(1 + \frac{z/a}{1 + |z/a|}\right)$$

$$\Phi = \frac{1}{a}\int_0^z \exp(-\pi(z/a)^2)\,\mathrm{d}z$$

where the value of a depends on the length of the gap, g. There is little point in giving further discussion of this type of approximation.

3.1.2 The value of ω

Values for ω were first calculated some 60 years ago [5] by three rather different methods which gave values of 1.318, 1.32 and 1.315. These values have appeared in many textbooks since then, including the first edition of the present work, frequently without specific attribution.

A value of 1.326 227 5, obtained by a numerical integration of equation (3.6), was quoted by Verster in 1963 [6], but this seems to have passed unnoticed and has not been cited in a textbook. This omission might suggest that there were doubts about the accuracy of this value and a recalculation would seem appropriate.

It is now quite practicable to evaluate the integral with much higher precision and accuracy, and we have used a standard series representation of $I_0(\kappa)$ [7] to determine values of the integrand from 0 to 50 at intervals of 0.001. We have then used three quadrature expressions, each at intervals of 0.01, 0.002 and 0.001, to calculate the integral. While there are differences beyond the 11th decimal place all the results are consistent with a value of $\omega = 1.326\,227\,505\,1$ in complete agreement with Verster, but with three more significant figures. This confirmation of Verster's value underlines the question of why an older, approximate, value was still being cited 25 years later [1, 2].

The numerical methods described in this chapter cannot calculate the field for a true zero gap lens, but values can be found for progressively smaller gaps and may be extrapolated to zero.

While it is no surprise that the Bessel function expansion method of section 3.6 gives an excellent agreement, two other methods have been applied to the problem and also agree well with our calculated value.

The finite difference method of section 3.5 lends itself to a simple type of extrapolation in which both the size of the gap and the density of the finite lattice are extrapolated together. With a cylinder radius, R_1, in terms of the lattice spacing, the smallest value of the gap to diameter ratio which can be modelled is $1/R_1$. Figure 3.2 shows values of the field at the centre of the gap as a function of $1/R_1^2$, which is proportional to the lattice density. A linear fit to these data extrapolates to 1.326 227 506 with a standard deviation of 5 in the last figure. This is a very satisfactory agreement. Similar agreement has been found by Read [8] using a special version of the CPO2D program which uses the boundary element method of section 3.2.

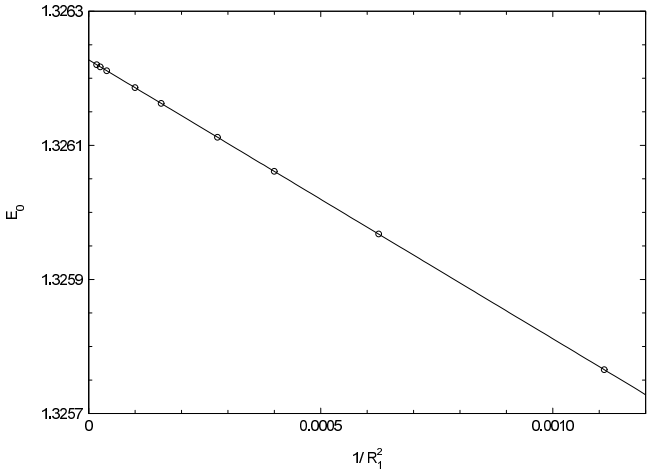

Figure 3.2 Values of the field at the centre of a two cylinder lens of radius R_1. Extrapolation to the zero value of $1/R_1^2$ implies zero gap length and also infinite lattice density.

3.2 The boundary element method

The solution of Laplace's equation requires the conditions to be known over a boundary which completely encloses the space. Almost without exception this calls for some approximation of the conditions in the gaps between the electrodes. The boundary element method, also known as the *charge density method*, does not have this constraint. It is in principle a very simple method: any electrode may be replaced by a system of charges at the electrode surface provided that the potentials they produce are everywhere the same as those produced by the electrode. In particular, the potential of the electrode itself must be reproduced by the charge distribution. The formal way to express this is

$$V_k = \frac{1}{4\pi\epsilon_0} \sum_{j=1}^{n} \int_{S_j} \frac{\sigma_j(\boldsymbol{r}_j)\,\mathrm{d}S_j}{|\boldsymbol{r}_j - \boldsymbol{r}_k|} \tag{3.11}$$

where V_k is the potential at a position $\boldsymbol{r}_k$ due to charges $q_j = \sigma_j\,\mathrm{d}S_j$ at positions $\boldsymbol{r}_j$ on each of n electrodes. In the present context of round lenses with axial symmetry, the problem reduces to the summation of the potentials at a point due to rings of charge on all the electrodes. The geometry of this situation is illustrated in figure 3.3 in which the potential at the ring of radius r_i due to the charge on the ring of radius r_j is to be found. In terms of the distances shown in this diagram the denominator of equation (3.11) is

$$l = [r_i^2 + r_j^2 - 2r_i r_j \cos(\alpha_i - \alpha_j) + (z_i - z_j)^2]^{1/2}.$$

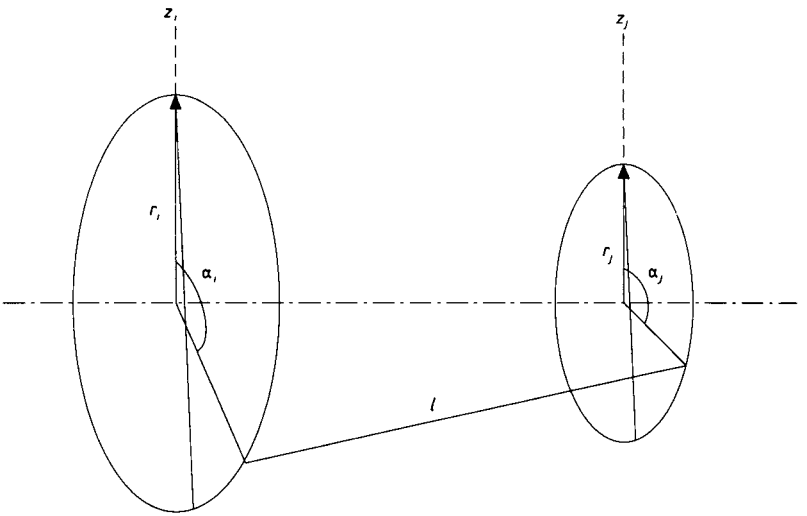

Figure 3.3 Diagram to indicate the coordinates appropriate to the calculation of the potential of a point due to a ring of charge.

The absolute values of the angles are arbitrary because of the axial symmetry so, for convenience, let $\alpha_i = \pi$ which allows us to simplify this expression as

$$\begin{aligned} l &= [(r_i + r_j)^2 - 4r_i r_j \sin^2(\alpha_j/2) + (z_i - z_j)^2]^{1/2} \\ &= \left\{[(r_i + r_j)^2 + (z_i - z_j)^2]\left(1 - \frac{4r_i r_j \sin^2(\alpha_j/2}{(r_i + r_j)^2 + (z_i - z_j)^2}\right)\right\}^{1/2}. \end{aligned}$$

The total charge on the ring, Q_j, can be written in terms of a line density, $\lambda_j = Q_j/2\pi r_j$, and each term in the summation of equation (3.11) becomes

$$\begin{aligned} &\frac{1}{4\pi\epsilon_0}\int_0^{2\pi} \frac{\lambda_j r_j \, \mathrm{d}\alpha_j}{l} \\ &\qquad = \frac{1}{\pi\epsilon_0} \frac{\lambda_j r_j}{[(r_i + r_j)^2 + (z_i - z_j)^2]^{1/2}} \\ &\qquad\qquad \times \int_0^{\pi/2} \left(1 - \frac{4r_i r_j \sin^2(\alpha_j/2)}{(r_i + r_j)^2 + (z_i + z_j)^2}\right)^{-1/2} \mathrm{d}(\alpha_j/2). \end{aligned}$$

The integral is just $K(\tau)$, the complete elliptic integral of the first kind,

$$\int_0^{\pi/2} (1 - \tau^2 \sin^2 \beta)^{-1/2} \, \mathrm{d}\beta$$

with

$$\tau^2 = \frac{4r_i r_j}{(r_i + r_j)^2 + (z_i - z_j)^2}$$

and so in terms of τ and Q_j the potential at a point i can be written as

$$V_i = \sum_{j=1}^{n} \frac{Q_j}{4\pi^2\epsilon_0} \frac{\tau}{\sqrt{r_i r_j}} K(\tau). \tag{3.12}$$

The practicability of solving this set of equations depends on a sensible choice of 'ring' elements and the distribution of charge between them. The charge density on the surface of an electrode is directly proportional to the electric field normal to the surface and a preliminary estimate of this field may be made in other ways, such as by the finite difference method. The charge density near the edge of an electrode will be high and change rapidly so small rings are needed here: conversely, deep inside a cylindrical electrode the charge density will be small and a substantial length of the cylinder may be used as a single 'ring'. This choice of elements is probably the factor which most severely limits the overall accuracy of the method. The most direct way of solving the problem is to treat equation (3.12) as a matrix equation relating the column vectors $\boldsymbol{V}_i$ and $\boldsymbol{Q}_j$ with the individual elements being functions of the positions of the two rings involved, but other methods have been used [9].

The use of discrete rings of charge leads to a non-uniformity of potential on the electrode surfaces. In place of a constant potential, the potential will vary in a 'saw-tooth' fashion with discontinuities at the edges of the ring elements. This effect can be minimized by making the charges on each element the same. An alternative approach would be to use a continuous charge distribution with a surface charge density described analytically.

Even though the only boundary conditions which need to be applied are those on the electrode surfaces, the effects of field penetration from charges on the surface of the vacuum system enclosing the electrodes should be assessed and the electrodes designed to minimize such effects.

Once the charge distribution consistent with the required electrode potentials has been found, the axial potential distribution follows readily, since the elliptic integrals are then just equal to 1. The gradient of the axial potential can also be expressed in closed form and the function $T(z)$ of the Picht equation follows immediately.

When properly optimized, the boundary element method is capable of very high accuracy. There are commercial programs available which use the boundary element method and, in some cases, a limited, trial version is available for download from the Web†.

† The CPO program written by RB Consultants Ltd can be downloaded from cpo.ph.man.ac.uk and the Optics program written by Dr Li Wang can be downloaded from www.jps.net/liwang/download.htm.

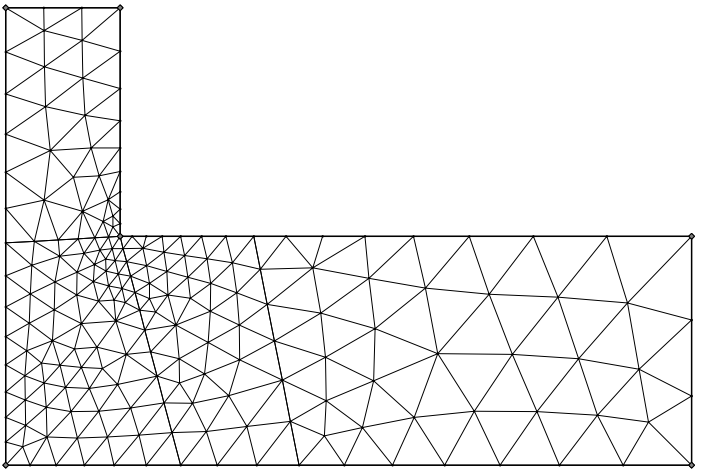

Figure 3.4 The right hand half of a two cylinder lens showing a pattern of triangular elements, suitable for analysis by the finite element method. Note that the sizes of the elements are far from constant.

3.3 The finite element method

Instead of treating the space within a system of electrodes as continuous, we may establish a set of discrete points at which we determine the potential by some means. The finite elements of this method are, usually, triangles and it is the potentials at the vertices of these triangles which we seek to determine.

Figure 3.4 shows a system of triangular elements appropriate to the solution of the potential distribution in a two cylinder lens with thick walls. We consider the walls to be thick compared to the length of the gap for two reasons. For purposes of calculation it is necessary to specify the potential at all boundaries, and at the outer edge of the gap a linear variation with position is usually a good approximation and one which can be tested. In addition, for practical purposes, the thick walls reduce the effect of field penetration from, for example, the walls of the vacuum chamber. This factor is important whatever method is used for the calculation.

Only half of the lens system is shown in figure 3.4 as the system is symmetric about the mid-plane. If we are representing a lens with potentials 1 and 0 on the left hand and right hand electrodes, we should set the potential at all points on the mid-plane to be 0.5 and at points at the outer edge of the gap to fall linearly from 0.5 to 0.

The axial potential gradient has a maximum value at the centre of the lens and is also large close to the inner end of the cylinder because of the 90° change of the boundary equipotential. In both regions we should use elements of small size though elsewhere, particularly deep inside the cylinder, they may be much larger.

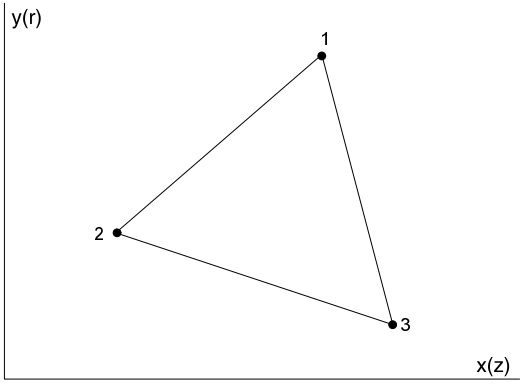

Figure 3.5 A representative element of the network illustrated in figure 3.4.

The physical principle behind the finite element method is that the potential distribution will be such that the potential energy of the electrostatic field is a minimum and so we need to express this potential energy in terms of the potentials at the vertices of each and all of the triangular elements.

3.3.1 The basic equations

We consider a typical element in figure 3.5 where we have shown both Cartesian and axisymmetric cylindrical coordinates. We shall use the latter for the lens example. We assume that the potential varies linearly with position and write

$$V_1 = V(z_1, r_1) = a + bz_1 + cr_1 \text{ etc} \tag{3.13}$$

where the coefficients can be expressed in terms of the coordinates of the vertices and their potentials.

$$\begin{aligned} a &= [V_1(z_2r_3 - z_3r_2) + V_2(z_3r_1 - z_1r_3) + V_3(z_1r_2 - z_2r_1)]/2S \\ b &= [V_1(r_2 - r_3) + V_2(r_3 - r_1) + V_3(r_1 - r_2)]/2S \\ c &= [V_1(z_3 - z_2) + V_2(z_1 - z_3) + V_3(z_2 - z_1)]/2S \\ S &= [z_1(r_2 - r_3) + z_2(r_3 - r_1) + z_3(r_1 - r_2)]/2 \end{aligned} \tag{3.14}$$

where S is the area of the triangle.

The potential energy of the electrostatic field is

$$W = \iint \Big[\Big(\frac{\partial V}{\partial r}\Big)^2 + \Big(\frac{\partial V}{\partial z}\Big)^2\Big]\frac{\epsilon_0}{2} 2\pi r \, \mathrm{d}r \, \mathrm{d}z. \tag{3.15}$$

The condition for minimum potential energy is that

$$\frac{\partial W}{\partial V_1} = \frac{\partial W}{\partial V_2} = \cdots = \frac{\partial W}{\partial V_k} = \cdots = \frac{\partial W}{\partial V_N} = 0. \tag{3.16}$$

For a single element, $\frac{\partial V}{\partial r} = c$ and $\frac{\partial V}{\partial z} = b$ so the associated potential energy can be written as

$$\begin{aligned} \Delta W &= \frac{\epsilon_0}{2}\big(b^2 + c^2\big) 2\pi r \, \mathrm{d}r \, \mathrm{d}z \\ &= \pi r \epsilon_0 S\big(b^2 + c^2\big). \end{aligned}$$

From figure 3.4 it is clear that any given node is common to a number of triangular elements and the potential at that node affects the potential energy of all of these. We should, therefore, write each term of equation (3.16) in the form

$$\frac{\partial W}{\partial V_k} = \frac{\partial}{\partial V_k}\Big(\sum_{j=k+1}^{k+n} \Delta W_j\Big) = 0 \tag{3.17}$$

where there are n elements having node k in common. We may move the differential operator inside the summation and write for a representative element

$$\frac{\partial(\Delta W)}{\partial V_j} = \pi r \epsilon_0 \big[b(r_{j+1} - r_{j+2}) + c(z_{j+2} - z_{j+1})\big].$$

The radial coordinate r in this equation has come from the volume element in equation (3.15) and should be assigned the value at the centroid of the triangle, that is the average of the values at the vertices. If we substitute for b and c using equation (3.14) this expression reduces to the sum of the potentials at the three nodes multiplied by coefficients which involve the coordinates of these nodes and which can be calculated once for all. Equation (3.17) can then be written as

$$\frac{\partial W}{\partial V_k} = \sum_{m=k}^{k+n} g_m V_m = 0.$$

This is one of a set of N equations which are usually solved by matrix methods. The solution is the set of free node potentials from which equipotential plots, such as that shown in figure 3.6, may be generated. Remember that all the nodes, including those at the boundaries, must be considered in evaluating these equations.

The finite element method is of limited accuracy for electrostatic lenses as the linear interpolations used in equation (3.13) militate against the determination

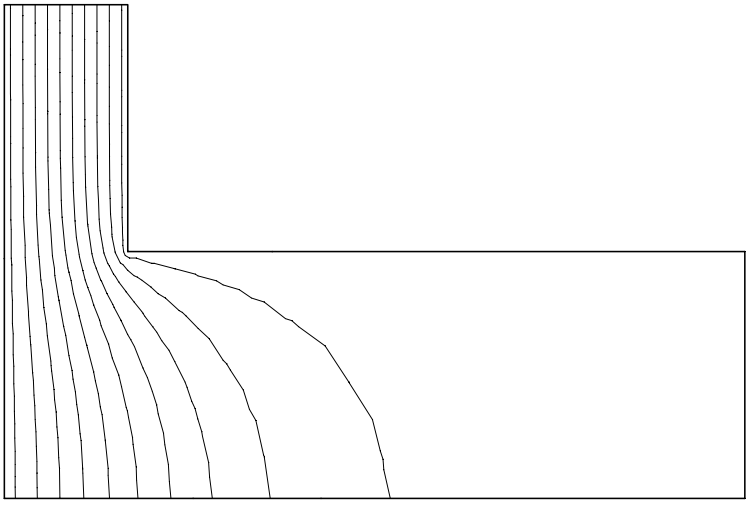

Figure 3.6 Equipotentials for the electrode and finite element system shown in figure 3.4. Notice that the equipotentials are not quite smooth; this is the result of using the linear potential variation of equation (3.13) within each triangular element.

of the accurate axial potentials and gradients required for ray tracing and the calculation of aberrations. It is a much more useful tool for the analysis of magnetic field problems and has been extensively used in this area.

Commercial programs based on the method exist, and the data of figures 3.4 and 3.6 were obtained using Student's QuickField 4.1† which is a limited version of a very powerful program. We noted earlier the need to vary the size of the elements such that, in regions of high field strength, the elements were small. QuickField allows the user to determine the sizes during the problem definition phase, but other programs, such as FlexPDE‡ will generate an optimal set of elements from the boundary conditions.

3.4 The finite difference method: relaxation

One of the simplest methods for calculating the potential distribution in a bounded region is to treat the space not as continuous, but as a regular lattice of discrete points. We can then replace the differential form of Laplace's equation by difference equations. We shall analyse systems of the

† This program, a development of one called ELCUT, can be downloaded from www.tor.ru/quickfield.

‡ FlexPDE lite can be downloaded from www.pdesolutions.com.

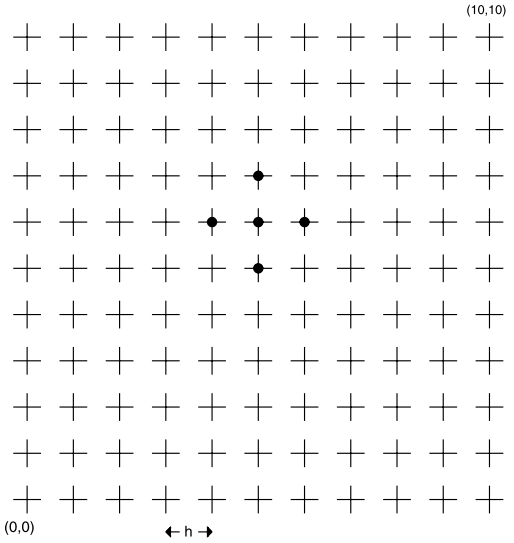

Figure 3.7 Lattice of points spaced by a distance h on which the potential may be relaxed using groups of five points (●).

cylindrical symmetry appropriate to electrostatic lenses later, but first consider a simpler case which allows us to illustrate some general features.

3.4.1 A two dimensional example

Laplace's equation in rectangular Cartesian coordinates is

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0. \tag{3.18}$$

We can represent this space by a rectangular array of points separated by distances h in both x and y directions as shown in figure 3.7, and identify five points each marked by the symbol ●. The point at the centre of this group is at (x, y). We can express the potential at the four outer points using a Taylor expansion of the potential at the centre point and its derivatives at that point.

$$\begin{aligned} V(x-h, y) &= V(x, y) - h\frac{\partial V}{\partial x} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial x^2} \\ &\quad - \frac{1}{6}h^3\frac{\partial^3 V}{\partial x^3} + \frac{1}{24}h^4\frac{\partial^4 V}{\partial x^4} - \cdots \\ V(x+h, y) &= V(x, y) + h\frac{\partial V}{\partial x} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial x^2} \\ &\quad + \frac{1}{6}h^3\frac{\partial^3 V}{\partial x^3} + \frac{1}{24}h^4\frac{\partial^4 V}{\partial x^4} + \cdots \end{aligned}$$

with two similar expressions, but involving y-derivatives, for $V(x, y - h)$ and $V(x, y + h)$. Adding all four equations we find

$$V(x-h, y) + V(x+h, y) + V(x, y-h) + V(x, y+h) - 4V(x, y) = h^2\left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2}\right)$$

ignoring higher order terms beginning with

$$+\frac{1}{12}h^4\left(\frac{\partial^4 V}{\partial x^4} + \frac{\partial^4 V}{\partial y^4}\right).$$

Equation (3.18) shows that the right hand side of this equation is zero. The potential at any point on the lattice can therefore be written as the mean of the potentials at the four nearest neighbour points. In order to apply this result to a real problem, fixed values of potential must be assigned to all the points on a boundary enclosing a region of our two dimensional space. For the non-boundary points it is sometimes possible to guess approximate potentials, but unless the array is quite large it is usually enough to set all the other potentials to zero. We then move systematically over the space replacing the potential at each point by the appropriate average. It is usual to start at a place on the boundary where regions at different potentials meet. As an example, suppose that the potentials at points $(x, 10)$ and $(0, y)$ are fixed at potential 1 and those at $(0, x)$ and $(10, y)$ are fixed at zero, x and y taking all values from 1 to 9; the potentials at the corners are set to 0.5, 1, 0.5, 0 in sequence, defining a square boundary. We would then begin by writing

$$\begin{aligned} V(1, 1) &= \tfrac{1}{4}[V(0, 1) + V(2, 1) + V(1, 0) + V(1, 2)] \\ &= \tfrac{1}{4}[1 + 0 + 0 + 0] = 0.25 \end{aligned}$$

then increase x to the end of the row and then move up one row, return to the left and so on. A single passage across the array is not going to give us the solution and we must repeat the whole cycle many times until the changes from one cycle to the next are sufficiently small. This process of repeated iteration is known as relaxation. Notice the change in notation here: distances are expressed in units of h.

3.4.1.1 Improving the speed and precision. The relaxation method outlined above will always give a solution, but the precision will depend on the scale of the array of points and the time taken to converge to an adequate degree will increase as the square of the number of points. This conflict between speed and precision can be resolved to some extent by first relaxing a rather coarse array and then inserting additional points between the array points, interpolating to find starting potentials for these new points and then continuing the relaxation process. However, a much greater increase in speed can be obtained by a procedure known

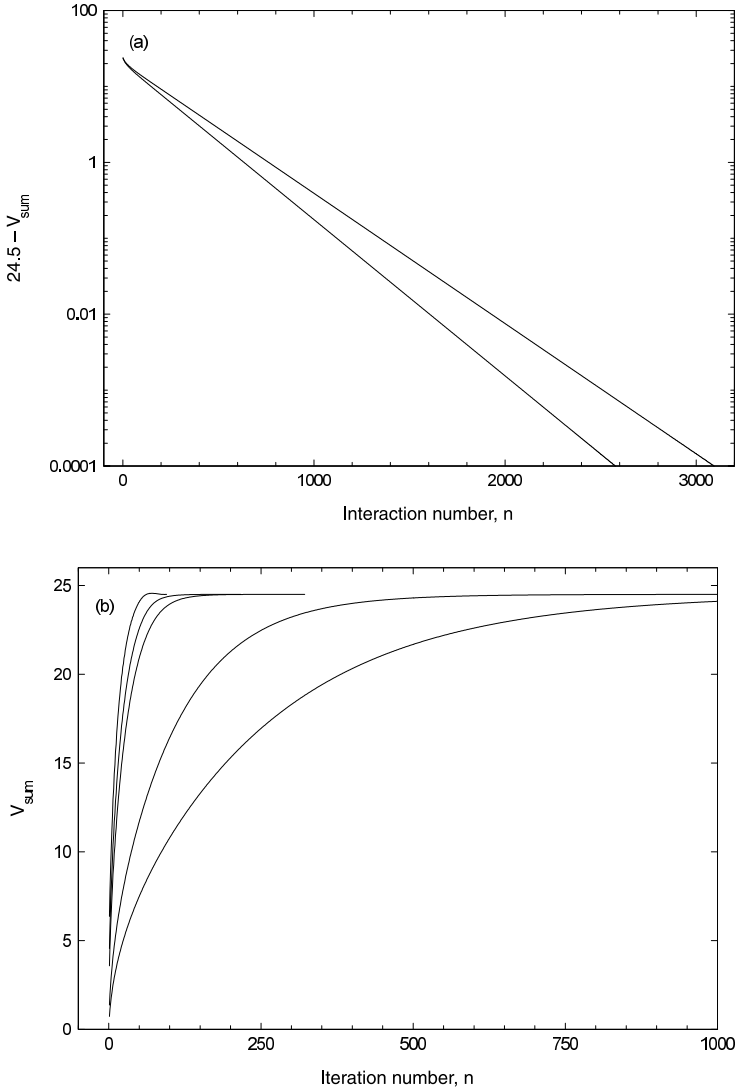

Figure 3.8 The approach of the potential on the diagonal of a square lattice (51×51 points) to the final value. (a) The difference between the potential after n cycles of iteration and the final potential, illustrating the exponential approach. The upper curve and those of figure 3.8(b) are the result of five point iteration; the lower curve is the result of nine point iteration, discussed later. (b) The potential after n cycles of iteration with acceleration parameters of (from the right) 1.0, 1.4, 1.8, 1.85, 1.90.

as 'successive over-relaxation'. The basis for this is that each time we calculate the potential at some point, we move it towards its final value, but by only a small amount, and indeed by an amount which is progressively smaller. Over-relaxation involves replacing the potential at a point not by the simple mean of that at its four neighbours but, recognizing that next time around it will be changed even more, we make a bigger change immediately. We write

$$V(x, y) = \frac{g}{4}(V(x-1, y) + V(x+1, y) + V(x, y-1) + V(x, y+1)) - (g-1)V(x, y)^{\dagger}$$

where $V(x, y)^{\dagger}$ is the value of $V(x, y)$ from the preceding iteration and g is called the acceleration parameter. As an example of the increased speed that can be obtained by appropriate choice of acceleration parameter, we show in figure 3.8 the behaviour of a square array similar to that of figure 3.7, but with 51 points on a side. We know from symmetry that the potential of all the points on the diagonal with $x = y$ will become 0.5 eventually and so the sum of the potentials for the 49 points with $0 < x = y < 50$ will be 24.5. Figure 3.8(a) shows the exponential approach to this final value for no acceleration, and figure 3.8(b) shows, on linear scales, the sum potential for a number of values of g. With no acceleration the sum reaches a value within 0.0001 of 24.5 only after more than 3000 cycles of iteration, while for $g = 1.90$ this condition is reached after 95 cycles. Notice that the final value is reached, in this case, *from above.* The effect of increasing g from 1 to larger values resembles very much the effect of reducing the damping of a resonant circuit and, if g is increased too much, the system becomes unstable and will oscillate for $g \geqslant 2$; it is the first signs of this which we see for $g = 1.90$. The maximum safe value of g depends on the size of the array and also on its shape. This example has only 2401 free points and would be regarded as a rather small array. We commonly use arrays of up to some 500 000 points with aspect ratios of six to 20 and have found no serious problems with values as high as 1.97. The Pascal program **RELAX51** used for this calculation is shown, in outline, on the following page, and a compiled version, with certain refinements, is included on the program disk.

3.4.1.2 Estimates of g. Expressions from which the value of the optimum value of the acceleration parameter may be estimated are given by Press *et al* [7] and figure 3.9 is based on these expressions. It is clear that only for very small arrays will values less than 1.7 be needed.

```
PROGRAM relax51;
VAR    V:array[0..50,0..50] of double;
       x,y,c:integer;
       g,Vsum:double;
PROCEDURE average;
VAR    x,y:integer;
BEGIN
  for x:=1 to 49 do BEGIN
    for y:=1 to 49 do BEGIN
    V[x,y]:=g*(V[x+1,y]+V[x-1,y]+V[x,y+1]+V[x,y-1])/4
           -(g-1)*V[x,y];
    END;
  END;
  Vsum:=0;
  for x:=1 to 49 do Vsum:=Vsum+V[x,x];
  writeln(c:4,Vsum:16:5);
END;
BEGIN
  REPEAT
    REPEAT
       clrscr;
       gotoxy(16,6);
       write('ENTER a value for the acceleration parameter,
             g: ');
       c:=wherex; writeln;
       gotoxy(16,7);
       write('(1',chr(243),' g < 2), but g=0 terminates
             the program');
       gotoxy(c,6);
       readln(g);
       IF g=0 THEN exit;
    UNTIL (g >= 1) and (g < 2);
    clrscr;
    c:=1;
    Vsum:=0;
    for x:=1 to 50 do for y:=0 to 49 do V[x,y]:=0;
    for x:=0 to 49 do V[x,50]:=1;
    for y:=1 to 49 do V[0,y]:=1;
    V[0,0]:=0.5;
    V[50,50]:=0.5;
    while abs(Vsum - 24.50) > 0.0001 do BEGIN
       average;
       c:=c+1;
    END;
    readln;
 UNTIL g=0;
 END.
```

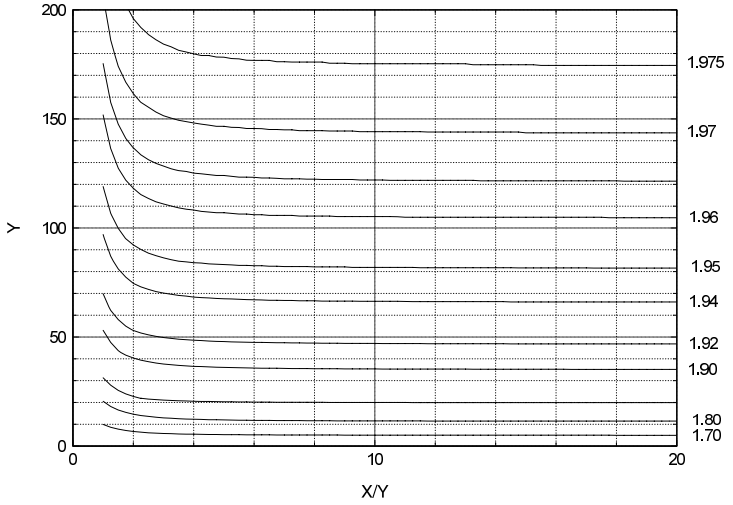

Figure 3.9 Estimates of the acceleration parameter, g, for arrays of $Y(R) \times X(Z)$ points. The abscissa is the aspect ratio, X/Y, which allows values for a wide range to be shown compactly.

3.4.2 Cylindrical symmetry

In cylindrical polar coordinates Laplace's equation (2.1) is

$$\frac{\partial^2 V}{\partial z^2} + \frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} = 0.$$

Consider a rectangular array of points separated by distances h with the polar axis forming the lower boundary. We can write for the potential close to the point (z, r)

$$V(z, r+h) = V(z, r) + h\frac{\partial V}{\partial r} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial r^2} + \frac{1}{6}h^3\frac{\partial^3 V}{\partial r^3} + \frac{1}{24}h^4\frac{\partial^4}{\partial r^4} + \cdots \tag{3.19a}$$

$$V(z+h, r) = V(z, r) + h\frac{\partial V}{\partial z} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial z^2} + \frac{1}{6}h^3\frac{\partial^3 V}{\partial z^3} + \frac{1}{24}h^4\frac{\partial^4}{\partial z^4} + \cdots \tag{3.19b}$$

with similar expressions for the two other neighbouring points. Taking appropriate sums and differences and ignoring terms of fourth order and above we can write

$$V(z+h,r)+V(z-h,r)-2V(z,r)=h^2\frac{\partial^2 V}{\partial z^2}$$
$$V(z,r+h)+V(z,r-h)-2V(z,r)=h^2\frac{\partial^2 V}{\partial r^2}$$
$$\frac{h}{2r}\big[V(z,r+h)-V(z,r-h)\big]=\frac{h^2}{r}\frac{\partial V}{\partial r}.$$

The right hand sides of these three equations are, term by term, equal to h^2 times the terms of Laplace's equation, and so the sum of the left hand sides must be zero. We can therefore write

$$\begin{aligned}V(z,r)&=\tfrac{1}{4}\big[V(z+h,r)+V(z-h,r)+V(z,r+h)+V(z,r-h)\big]\\&\quad+\frac{h}{8r}\big[V(z,r+h)-V(z,r-h)\big]\\&=\Bigg[V(z+h,r)+V(z-h,r)+\left(1+\frac{h}{2r}\right)V(z,r+h)\\&\quad+\left(1-\frac{h}{2r}\right)V(z,r-h)\Bigg]\Big/4\end{aligned}$$

where the latter version is better suited to computation as the r-dependent coefficients can be calculated once for all.

These expressions cannot be applied to points on the axis, as the final term would become infinite. For this case we have to consider six points: two axial points, one to either side of the target point, and four points in a plane perpendicular to the axis.

$$V(z,0)=\tfrac{1}{6}\big[V(z+h,0)+V(z-h,0)+4V(z,h)\big].$$

As in the two dimensional case, the convergence of the calculation can be made significantly faster by over-relaxation.

$$\begin{aligned}V(z,r)&=\frac{g}{4}\Bigg[V(z+h,r)+V(z-h,r)+\left(1+\frac{h}{2r}\right)V(z,r+h)\\&\quad+\left(1-\frac{h}{2r}\right)V(z,r-h)\Bigg]-(g-1)V(z,r)^{\dagger}\\V(z,0)&=\frac{g}{6}(V(z+h,0)+V(z-h,0)+4V(z,1))-(g-1)V(z,0)^{\dagger}.\end{aligned}$$

3.4.2.1 Boundary conditions. It is a straightforward matter to apply the conditions at the electrode surfaces, but the gaps between electrodes can present problems and there are other 'boundaries of convenience' which can be introduced to reduce the computational space and time requirements. Considerable reductions

can be made by recognizing planes of symmetry in the problem. Figure 3.10 shows the electrodes of two aperture lenses, one with two apertures and the other with three. The origin of coordinates is at the centre of the lens. In the case of the two aperture lens, the symmetry is obvious and a suitable set of boundary conditions would be $V = 0$ on the left hand electrode, $V = 0.5$ on the plane of symmetry (a boundary of convenience) and a linear rise from 0 to 0.5 along the line from a to b. It is easy to test the validity of this last condition by applying it at a large radial distance and examining the change in potential distribution between the aperture plates as a function of radial position: it is usually very good provided that $(D_3 - D_1)/A$ is greater than about two. Only the mesh in the region indicated needs to be relaxed. The potential gradients towards the left hand end of this region are quite low and the relaxation process will converge very slowly if it is started at that end. It is far better to start the process in a region which will have large potential gradients such as the junction between the mid-plane boundary and the outer cylindrical surface as this will tend to drive the effects of the different boundaries across the mesh. One might think of the method as a type of diffusion process and, in a later section, we shall see that this is quite a realistic description.

For the three aperture lens it is necessary to find the potential distribution for two cases. The first, asymmetric, case has potentials of 0, 0.5 and 1 on the left, centre and right electrode respectively, and a potential of 0.5 can be assigned to the centre plane leaving, again, only the hatched area to be analysed. The symmetric case, with potentials 0, 0.5 and 0 on the three electrodes needs more thought because the potential on the centre plane is not part of the fixed boundary conditions but has to be found. This is not difficult. For $0 < r < D_1/2$ we write

$$V(0, r) = \tfrac{1}{4}[V(0, r - h) + V(0, r + h) + 2V(-h, r)] + \frac{h}{8r}[V(0, r + h) - V(0, r - h)]$$

and on the axis

$$V(0, 0) = \tfrac{1}{3}[V(h, 0) + 2V(0, h)]$$

with appropriate modifications for acceleration.

If the electrode potentials are set at V_1, V_2 and V_3, the axial potential can be written as

$$V(z) = V_1 + (V_3 - V_1)\Phi_a(z) + (2V_2 - V_1 - V_3)\Phi_s(z)$$

where $\Phi_a(z)$ and $\Phi_s(z)$ are the axial potential functions found for the asymmetric and symmetric cases respectively.

3.4.3 Target conditions

For the example of a square mesh considered in section 3.4.1, the values of the potential on all the diagonal points were known from the symmetry, and testing for convergence by examining the difference between the value after each iteration

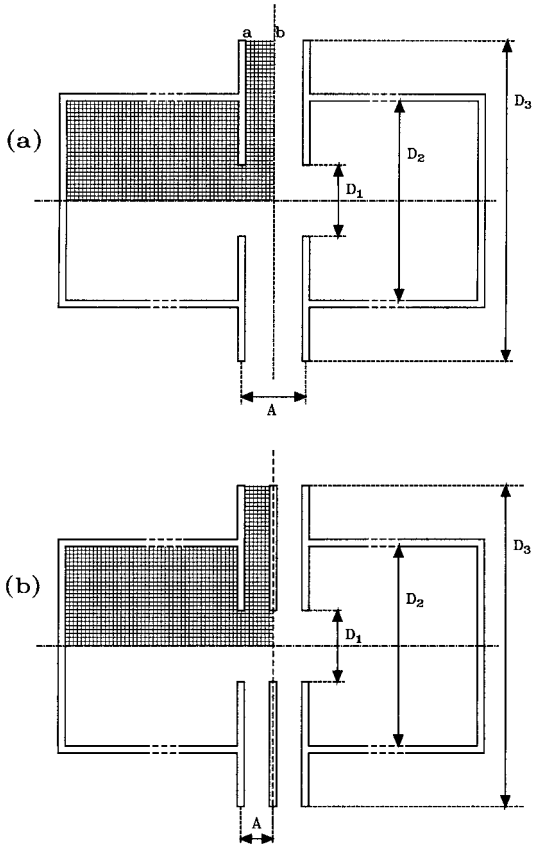

Figure 3.10 Aperture lenses of two and three elements showing the important dimensions and the regions over which the potential must be solved. The cylinders have to be bounded by a known potential at their outer ends, but the length would normally be much greater than is shown here, as is implied by the broken-line sections.

cycle and this known value was not a problem. Even in this situation, a simple test may fail if the acceleration parameter is too large, because the overshoot illustrated in figure 3.8(b) for $g = 1.90$ may, by chance, satisfy the condition on the rising part. In general the true potential distribution is not known *a priori* and some other method of testing for the convergence of the iterative procedure has to be employed.

It is a fair assumption that *eventually* the potentials at all points will be as close to the true values as the discrete nature of the analysis will allow. At some points the potential will be too high and at others it will be too low, but after the next cycle of iteration these specific values will change and may, indeed, alternate on subsequent cycles. The magnitude of these changes will be some measure of the

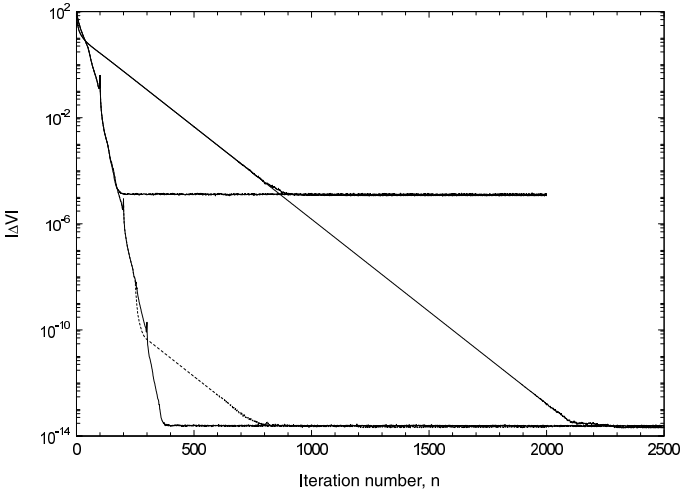

Figure 3.11 The dependence of $|\Delta V|$ on the iteration number for the array of **RELAX51** calculated using the five point method using single and double precision arithmetic and with two values of the acceleration parameter, g.

effect of the discrete nature of the process and will also depend on the acceleration parameter. It is useful to examine their behaviour as it may be a guide to the completion of the iterative process.

It would be impracticable to examine the changes at every point for a large array, typical of lens systems, but we shall see later that all the lens parameters of general interest can be deduced from the axial potential and changes in this can be an adequate guide. We write $\delta V(j)$ as the change in the potential at the jth (axial) point, and form the sums

$$\Delta V = \sum_j \delta V(j) \qquad |\Delta V| = \sum_j |\delta V(j)|.$$

We expect the first sum to take either sign and to approach zero, but the second sum is always positive and will decrease, tending to some finite value. This limiting value will depend on the precision of the calculation and also, though this is only apparent for quite large arrays, on the acceleration parameter, g.

Figure 3.11 illustrates the behaviour of $|\Delta V|$ for the array of **RELAX51** where the δV are summed along the diagonal and the iteration is continued for 2000 or more steps. The results are shown for two values of g, 1.60 and 1.90, using single and double precision arithmetic, and also for a sequence, shown by the broken lines, in which $g = 1.90$ until $|\Delta V| < 10^{-8}$ from which point the lower value of g is used.

There are a number of points of interest in this figure. The asymptotic values for the single and double precision calculations do not depend significantly on the

value of g. We shall see later that there is a much clearer dependence in the case of large arrays. The approach to the asymptotes has a slope which depends only on the value of g even in the case where most of the approach is made more rapidly, using a larger initial value. This shows that much calculation time can be saved by starting the iteration with a large, even *excessive* value. Again, this will be discussed further in the context of lens geometries in the appendix. Finally, there are glitches separated by 100 iterations in the approach of the $g = 1.90$ curve to its asymptote and, at the point where the value of g is reduced, the value of $|\Delta V|$ falls rapidly before adopting the slower decrease appropriate to the new value.

The demonstration program **RELAXCYL.EXE** on the disk uses this criterion and displays both ΔV and $|\Delta V|$ at each step and halts when $|\Delta V| = 0.00001$. Note that the sign of ΔV alternates to some extent for large values of g. The glitches in figure 3.11 mentioned above correspond to large negative values of ΔV. The reader may care to examine the dependence on g of the number of iteration cycles required to reach the end point. A graph of $\log n$ against g emphasizes the importance of the correct choice of value, while a graph of n against $\log(2 - g)$ is also interesting. Both graphs show the effects of incipient oscillation for values above the optimum.

3.4.4 Precision and accuracy

The precision of the final result of a relaxation calculation depends on the size of the array. With other factors being the same, the precision increases directly with the number of points in the array. The *accuracy* is also affected, because the contribution of the neglected higher order terms of the Taylor expansion becomes smaller.

There are particular problems close to edges and corners of electrodes, where the potential changes rapidly. It is possible to alleviate these if the mesh density is increased locally in these regions by reducing h to a half or a quarter of its value over the main system [10]. Unfortunately, it is just at the edges of electrodes that the boundary conditions are not well known and increasing the mesh density may give a false sense of accuracy when only the *precision* of the solution to a somewhat different problem has been improved.

3.4.4.1 The limitations of the five point method. The accuracy of the five point method may be estimated by noting that the potential at a given point differs from the mean of that at the nearest neighbours by an amount proportional to h^4 or N^{-2} where N is the number of mesh points. However this is not the absolute accuracy, because these latter points do not have their 'correct' values either so there is a progressive change which will depend on N, leaving the absolute accuracy proportional to N^{-1}. It is very easy to exploit this dependence by solving the same problem with different mesh spacing. We noted earlier that the speed of calculation could be increased by repeated halvings of the mesh separation, and if the problem is tackled in this way, the information needed to *extrapolate* to an infinite number of points is available. In figure 3.12, values of the potential at five points on the axis of a two cylinder lens calculated with mesh spacings of $D/40$, $D/80$ and $D/160$

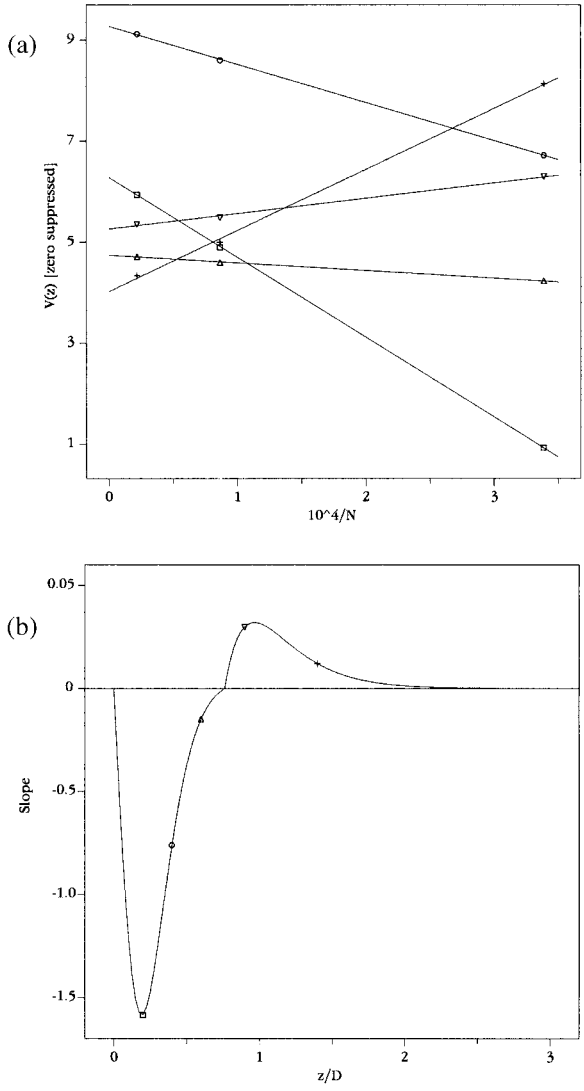

Figure 3.12 (a) The potential at five points on the axis of a cylindrical lens calculated by the five point method on lattices having spacing of $D/40$, $D/80$ and $D/160$ plotted against the reciprocal of the number of lattice points, illustrating extrapolation to an infinite lattice density. The five sets of data are shown with substantially suppressed zeros to emphasize the variation in slope. The data for 0.9 and 1.4 diameters are shown on ordinate scales magnified by 10 and 100 respectively. (b) The dependence of the slope of the lines of (a) (identified by their symbols), and similar plots, on axial position. The scale is magnified by a factor of ten for the positive values.

are plotted against the reciprocal of the number of mesh points. There is a clear dependence on N^{-1} though the slope varies and even changes sign. While this extrapolation can only be applied at the spacing of the coarsest mesh, the smooth variation of the slope can itself be interpolated so that the values of the potential at intermediate points can be corrected.

This position dependent slope implies that the five point method can never give the true axial potential and it is only this extrapolation procedure which allows the true potential to be found.

3.4.4.2 The nine point method. It is possible to increase the accuracy of the calculation by considering not just the average of the potentials at four nearest neighbours, but by taking account of four further points. This leads to relaxation methods known as 'nine point relaxation' in distinction to the 'five point relaxation' we have discussed so far. If we consider the four points a distance $\pm 2h$ from the target point in the two dimensional array [11] we can write

$$V(x-2, y) + V(x+2, y) + V(x, y-2) + V(x, y+2) - 4V(x, y) = 16h^2\left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2}\right)$$

and this time the first neglected term is

$$+\frac{4}{3}h^4\left(\frac{\partial^4 V}{\partial x^4} + \frac{\partial^4 V}{\partial y^4}\right)$$

which is 16 times as great as for the four nearest neighbour case. We can eliminate this term, leaving the sixth order term as the first to be neglected, if we give the nearest neighbour sum a weight of 16 and the '$2h$' sum a weight of -1

$$\begin{aligned} 60V(x, y) = {} & 16[V(x-h, y) + V(x+h, y) + V(x, y-h) + V(x, y+h)] \\ & - [V(x-2h, y) + V(x+2h, y) + V(x, y-2h) + V(x, y+2h)]. \end{aligned}$$

This choice of additional points leads to some problems close to boundary points and it is more usual to use the points at the corners of a square to eliminate the fourth order term. The expansion of the potential at these points to fourth order requires 15 terms:

$$\begin{aligned} V(x+h, y+h) = {} & V(x, y) \\ & + h\frac{\partial V}{\partial x} + h\frac{\partial V}{\partial y} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial x^2} + h^2\frac{\partial^2 V}{\partial x \partial y} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial y^2} \\ & + \frac{1}{6}h^3\frac{\partial^3 V}{\partial x^3} + \frac{1}{2}h^3\frac{\partial^3 V}{\partial x^2 \partial y} + \frac{1}{2}h^3\frac{\partial^3 V}{\partial x \partial y^2} + \frac{1}{6}h^3\frac{\partial^3 V}{\partial y^3} \\ & + \frac{1}{24}h^4\frac{\partial^4 V}{\partial x^4} + \frac{1}{6}h^4\frac{\partial^4 V}{\partial x^3 \partial y} + \frac{1}{4}h^4\frac{\partial^4 V}{\partial x^2 \partial y^2} \\ & + \frac{1}{6}h^4\frac{\partial^4 V}{\partial x \partial y^3} + \frac{1}{24}h^4\frac{\partial^4 V}{\partial y^4}. \end{aligned} \tag{3.20}$$

We leave it as an exercise for the reader to show that the fourth order terms can be removed by giving the nearest neighbour terms a weight of four and the corner terms a weight of unity.

The lower of the two curves of figure 3.8(a) is the result of this form of the nine point method. It reaches the target condition after 2577 cycles of iteration compared with the 3094 cycles taken by the five point method. The total calculation time is greater, however, being 0.35 m s per cycle compared with 0.7 m s per cycle on a 450 MHz Pentium III machine.

3.5 The nine point method applied to cylindrical geometry

The increases in precision and, especially, accuracy of the nine point method are so significant that a full derivation of the formulæ is appropriate.

In cylindrical polar coordinates, equation (3.20) becomes

$$\begin{aligned} V(z+h, r+h) = V(z,r) &+ h\frac{\partial V}{\partial z} + h\frac{\partial V}{\partial r} \\ &+ \frac{1}{2}h^2\frac{\partial^2 V}{\partial z^2} + h^2\frac{\partial^2 V}{\partial z \partial r} + \frac{1}{2}h^2\frac{\partial^2 V}{\partial r^2} \\ &+ \frac{1}{6}h^3\frac{\partial^3 V}{\partial z^3} + \frac{1}{2}h^3\frac{\partial^3 V}{\partial z^2 \partial r} + \frac{1}{2}h^3\frac{\partial^3 V}{\partial z \partial r^2} + \frac{1}{6}h^3\frac{\partial^3 V}{\partial r^3} \\ &+ \frac{1}{24}h^4\frac{\partial^4 V}{\partial z^4} + \frac{1}{6}h^4\frac{\partial^4 V}{\partial z^3 \partial r} + \frac{1}{4}h^4\frac{\partial^4 V}{\partial z^2 \partial r^2} \\ &+ \frac{1}{6}h^4\frac{\partial^4 V}{\partial z \partial r^3} + \frac{1}{24}h^4\frac{\partial^4 V}{\partial r^4} \end{aligned} \tag{3.21}$$

with three further expressions which differ from this only in the signs of the terms involving the odd order differentials.

3.5.1 The nine point formula for a general point

Laplace's equation relates only first and second derivatives and, if we are to improve the calculations beyond those considered in section 3.4.2, we have to eliminate the higher order derivatives from the eight expressions of equations (3.19) and (3.21). By taking appropriate sums and differences of these expressions, we can write explicit expressions for the higher order derivatives in terms of the potentials and the lower order derivatives.

$$
\begin{aligned}
h^3\frac{\partial^3 V}{\partial r^3} &= 3\big[V(z,r+h)-V(z,r-h)\big]-6h\frac{\partial V}{\partial r}\\
h^4\frac{\partial^4 V}{\partial r^4} &= 12\big[V(z,r+h)+V(z,r-h)\big]-24V(z,r)-12h^2\frac{\partial^2 V}{\partial r^2}\\
h^3\frac{\partial^3 V}{\partial r\partial z^2} &= \frac{1}{2}\big[V(z+h,r+h)+V(z-h,r+h)\\
&\qquad -\big(V(z+h,r-h)+V(z-h,r-h)\big)\big]\\
&\qquad +V(z,r-h)-V(z,r+h) \qquad (3.22)\\
h^4\frac{\partial^4 V}{\partial r^2\partial z^2} &= 4V(z,r)+V(z+h,r+h)+V(z-h,r+h)\\
&\qquad +V(z+h,r-h)+V(z-h,r-h)\\
&\qquad -2\big(V(z+h,r)+V(z-h,r)+V(z,r+1)+V(z,r-h)\big)\\
h^4\frac{\partial^4 V}{\partial z^4} &= 12\big(V(z+h,r)+V(z-h,r)\big)-24V(z,r)-12h^2\frac{\partial^2 V}{\partial z^2}.
\end{aligned}
$$

We can find relations between these higher differentials by differentiating Laplace's equation once and twice with respect to r and twice with respect to z.

$$
\begin{aligned}
\frac{\partial^3 V}{\partial r^3}+\frac{\partial^3 V}{\partial r\partial z^2}+\frac{1}{r}\frac{\partial^2 V}{\partial r^2}-\frac{1}{r^2}\frac{\partial V}{\partial r} &= 0\\
\frac{\partial^4 V}{\partial r^4}+\frac{\partial^4 V}{\partial r^2\partial z^2}+\frac{1}{r}\frac{\partial^3 V}{\partial r^3}-\frac{2}{r^2}\frac{\partial^2 V}{\partial r^2}+\frac{2}{r^3}\frac{\partial V}{\partial r} &= 0\\
\frac{\partial^4 V}{\partial r^2\partial z^2}+\frac{\partial^4 V}{\partial z^4}+\frac{1}{r}\frac{\partial^3 V}{\partial r\partial z^2} &= 0.
\end{aligned}
$$

Substitution of the five expressions (3.22) into these three equations leads to a single equation for $V(z,r)$ in terms of the potentials at the eight adjacent points. The use of a computer algebra package, in our case DERIVE†, is very useful at this stage.

$$
\begin{aligned}
60(8r^2+3h^2)V(z,r) &= 6(16r^2+7h^2)(V(z+h,r)+V(z-h,r))\\
&\quad +(96r^2+48rh+30h^2+23h^3/r)V(z,r+h)\\
&\quad +(96r^2-48rh+30h^2-23h^3/r)V(z,r-h)\\
&\quad +(24r^2+12rh+9h^2+13h^3/2r)(V(z-h,r+h)+V(z+h,r+h))\\
&\quad +(24r^2-12rh+9h^2-13h^3/2r)(V(z-h,r-h)+V(z+h,r-h)).
\end{aligned}
\qquad (3.23a)
$$

Again, this cannot be applied to points on the axis, but the calculation in terms of the potentials at the five nearest points in the plane is straightforward.

† Soft Warehouse, Inc., 3660 Waialae Avenue, Suite 304, Honolulu, HI, USA.

3.5.2 The nine point formula for an axial point

A Taylor expansion of the potential near the axis gives

$$V(z,h) = V(z,0) - \frac{h^2}{4}\frac{\partial^2 V}{\partial z^2} + \frac{h^4}{64}\frac{\partial^4 V}{\partial z^4} \tag{3.24a}$$

and the potential at the adjacent axial points follows from equation (3.19*b*) with $r = 0$ and the corresponding expression for $V(z-h,0)$.

$$V(z-h,0) + V(z+h,0) = 2V(z,0) + h^2\frac{\partial^2 V}{\partial z^2} + \frac{h^4}{12}\frac{\partial^4 V}{\partial z^4}. \tag{3.24b}$$

The potentials at the remaining two points are found by writing equation (3.19*b*) as $V(z \pm h, h)$, remembering that $\partial^2 V/\partial z^2$ has to be evaluated at these values of z :

$$\begin{aligned}
V(z-h,h) + V(z+h,h) &= V(z-h,0) + V(z+h,0) \\
&\quad - \left(\frac{h^2}{4}\frac{\partial^2 V}{\partial z^2}\right)_{z-h} - \left(\frac{h^2}{4}\frac{\partial^2 V}{\partial z^2}\right)_{z+h} + \frac{h^4}{32}\frac{\partial^4 V}{\partial z^4} \\
&= V(z-h,0) + V(z+h,0) \\
&\quad - \left(\frac{h^2}{2}\frac{\partial^2 V}{\partial z^2}\right)_{z} - \frac{h^4}{4}\frac{\partial^4 V}{\partial z^4} + \frac{h^4}{32}\frac{\partial^4 V}{\partial z^4} \\
&= 2V(z,0) + h^2\frac{\partial^2 V}{\partial z^2} \\
&\quad + \frac{h^4}{12}\frac{\partial^4 V}{\partial z^4} - \frac{h^2}{2}\frac{\partial^2 V}{\partial z^2} - \frac{h^4}{4}\frac{\partial^4 V}{\partial z^4} + \frac{h^4}{32}\frac{\partial^4 V}{\partial z^4} \\
&= 2V(z,0) + \frac{h^2}{2}\frac{\partial^2 V}{\partial z^2} - \frac{13}{96}\frac{\partial^4 V}{\partial z^4}.
\end{aligned} \tag{3.24c}$$

The differential terms in equations (3.24*a*–*c*) can be eliminated by taking the sum, weighted by multipliers of 34, five and seven, respectively, which gives the following expression for the potential at a point on the axis.

$$\begin{aligned}
58V(z,0) = 34V(z,h) &+ 5(V(z-h,0) + V(z+h,0)) \\
&+ 7(V(z-h,h) + V(z+h,h)).
\end{aligned} \tag{3.23b}$$

Equations (3.23) were first derived by Durand [12], but a more recent, and perhaps more accessible, derivation by Szilagyi [2], has a misprint in the coefficients of the $V(z, r \pm h)$ terms. Kasper [13] quotes an expression as valid for fields of various symmetries. However, for the axisymmetric case we are considering, his expressions give the coefficient of the $V(z, r+h)$ term as exactly four times that of the $V(z \pm h, r \pm h)$ terms which is only true in the limit of large r. The correct expression has been found by Edwards [10], but is not given in his

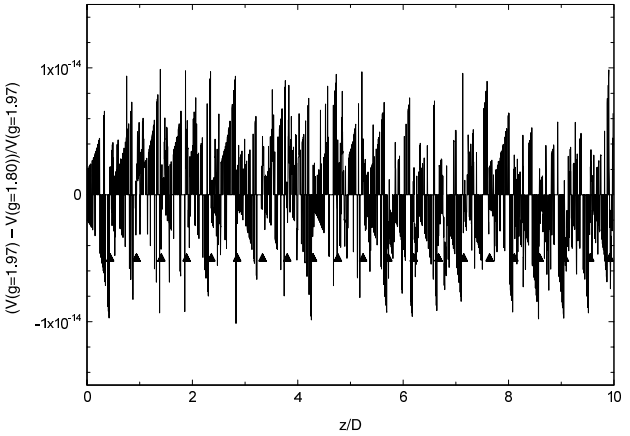

Figure 3.13 The fractional difference between the axial potential after sufficient cycles of iteration to ensure convergence. The two data sets are for 40 000 iterations with $g = 1.80$ and 10 000 iterations with $g = 1.97$.

paper, but only in a separately published appendix with only the barest outline of derivation.

3.5.3 The convergence of the potential in the lens geometry

The criterion that $|\Delta V|$ has reached some chosen value, or that it has reached its asymptotic value, is often adequate for a 'closed box' situation, but is not good enough in the context of lens design, where the values of the potential are not known and the range of these is very large. In this situation we must recognize that reaching the asymptotic value is only the beginning and subsequent cycles of iteration are required to drive the correct values of the potential along the length of the lens. The potential distribution has therefore to be tested for consistency under continued iteration. If the iteration has converged completely, values of the potential at each axial point will not have changed, to within the numerical precision, after some further cycles.

The most sensitive way to examine this is to calculate the *fractional difference* between the values

$$\frac{V_1(z) - V_2(z)}{V_2(z)}$$

for all values of the axial position z, where V_1 and V_2 correspond to different numbers of iterations. As an indication of the precision we can expect, figure 3.13 shows the comparison between two data sets after full convergence has been obtained. There is a clear periodicity which shows that we are seeing differences in the last significant digit of the double precision numbers we have used. The value of the potential at $z/D = 0$ is 0.5 while at $z/D = 9.9$ it is $1.0 \cdots \times 10^{-21}$. Each

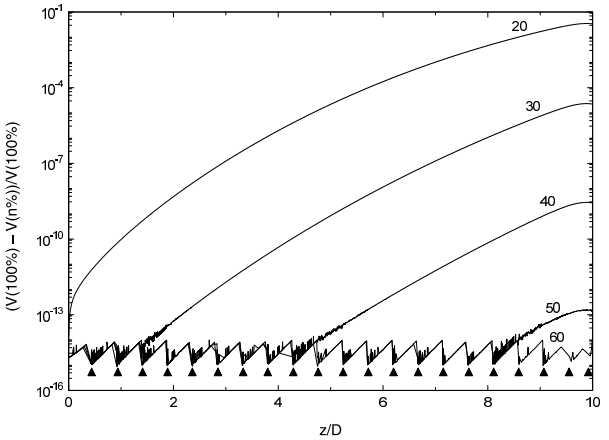

Figure 3.14 The fractional difference between the axial potential after 10 000 cycles of iteration with $g = 1.95$ and at five intermediate stages.

of the 21 ▲ marks the axial position where the number changes from 1.0×10^{-n} to $9.9 \times 10^{-(n+1)}$ and they echo the periodicity of the data. A histogram of these data with bins of width 10^{-15} shows that 48% are in the zero bin with almost equal numbers greater (26.3%) and less (25.7%) than zero.

Figure 3.14 shows the situation before convergence. The values of the fractional difference are so large that we cannot use the linear display of figure 3.13. The logarithmic display we have to use loses the negative data, but the progress of convergence along the axis is quite clear. After 20% of the final number of iterations the axial potential is nowhere correct; after 30% it has converged to $z/D \approx 1.2$; but it is only after some 55% that there is full convergence. In the calculations used in the preparation of the potential data file, **LPOTS.DAT**, for **LENSYS2.EXE** the tests for convergence were of this type, but were automated, as described in the Appendix, and were made at rather closer intervals.

3.5.4 Comparisons of the five and nine point methods

The advances in personal computer technology since the first edition have reduced the importance of the extrapolation procedure of section 3.4.4.1 and it is now practicable to apply the finite difference method to significantly larger arrays and to realize the advantage in accuracy of the nine point relaxation method. For the calculation of the potentials and fields used in the **LENSYS2.EXE** program we have used arrays of up to 240 by 2400 and 160 by 3200 points. The details of these calculations are discussed in the appendix, but it will be useful to compare some aspects of the five and nine point methods here.

To illustrate the similar precision of the two methods, we show in figure 3.15 the iteration of a cylindrical array system with 80 radial points and 1600 axial points for 8000 cycles. The iterations were started with $g = 1.95$ which was reduced to

TSMC EX. 1016 Page 61 of 139

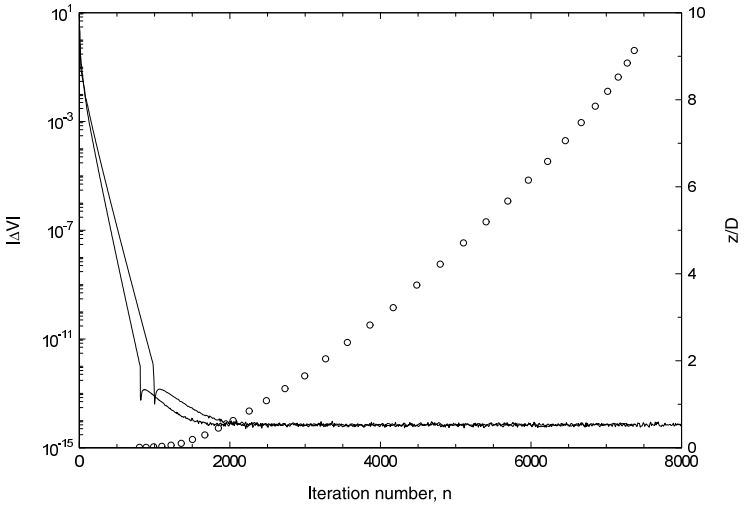

Figure 3.15 The variation of $|\Delta V|$ for a cylinder lens lattice having 80 radial and 1600 axial points calculated by both the five point (the upper curve) and nine point methods. The approach of the axial potential to its final value is indicated by the ○ curve and was determined by an extension of the procedure mentioned in section 3.5.3 and described in the appendix.

1.80 when $|\Delta V|$ fell below 10^{-12}. Both sets of data approach the same limiting value, 7.0×10^{-15} with standard deviations of 6.5×10^{-16}. The iteration times on a 120 MHz Pentium computer were 0.19 and 0.28 s per cycle for the 1000 and 822 cycles taken to reach the abrupt 'step' similar to that noted in the discussion of figure 3.11. The progress of the convergence of the axial potential is shown for the nine point case. There is little point in continuing the convergence test beyond $z = 9D$ because the potential is distorted by the zero potential 'end cap' at $z = 10D$.

The difference in accuracy is illustrated in figure 3.16. This compares the axial potential found after adequate convergence of both calculations. The figure is a comparison of the fractional difference between the potentials determined by the two methods, both with $g = 1.80$.

$$\frac{V_9(z) - V_5(z)}{V_9(z)}$$

where the notation should be transparent.

We know that the five point method does not converge to the correct potential and we shall see later that the nine point method does converge correctly in a certain sense, so the differences shown in this figure indicate the scale of the error in the five point method as applied to a lens geometry. The information in this figure is similar to that in figure 3.12.

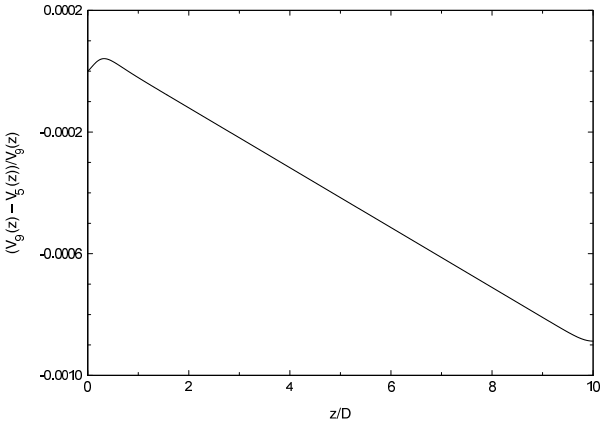

Figure 3.16 The fractional difference between the axial potentials of the lens of figure 3.15 calculated by the five point and nine point methods.

The finite difference method is one of the easiest to program, provided that the geometry is simple; curved surfaces need to be approximated by chords. It can be much more accurate than practical lens systems justify, giving very good focal and aberration data for both cylinder and aperture lenses. There are commercial programs available. The best known of these is **SIMION**†, which we have used to prepare figures 1.2(b), 4.3 and 5.6, but they all use the five point method which as we have seen has limitations which may be significant.

3.6 Bessel function expansions

We saw in section 3.1 that the potential in an axisymmetric situation could be expressed as

$$V(z, r) = (A_k e^{kz} + B_k e^{-kz}) J_0(kr) + V_0. \tag{3.1}$$

The general solution for a particular lens geometry will consist of a sum of such terms and, in contrast to the treatment of section 3.1, we choose to treat positive and negative values of kz explicitly. Figure 3.17 shows the geometry of a lens formed by two coaxial cylinders of equal diameter, D, separated by an axial distance, g, and having potentials V_1 and V_2. We show the walls of the cylinder to be of finite thickness as we shall later use a result from a finite difference calculation to resolve a problem inherent in this method. Three regions are indicated, in each of which we can write a specific expression for the potential. In regions I and III, within the

† A demonstration version of **SIMION** can be downloaded from www.srv.net/∼klack/simion.html.

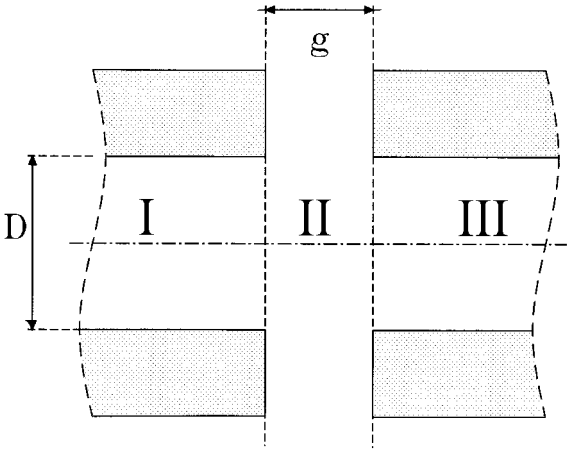

Figure 3.17 Schematic diagram of a two cylinder lens showing the dimensions and identifying the regions to which different expressions for the potential apply. The origin of the coordinates is at the centre of the system.

cylinders, only one of the terms of equation (3.1) has physical meaning; the other would become infinite as $|z| \to \infty$.

$$V_I(z,r) = V_1 + \sum_{n=1}^{\infty} A_n \exp(k_n z) J_0(k_n r) \tag{3.25a}$$

$$V_{II}(z,r) = \frac{V_1 + V_2}{2} + \left(\frac{V_2 - V_1}{g}\right) z + \sum_{n=1}^{\infty} [B_n \exp(-k_n z) + B'_n \exp(k_n z)] J_0(k_n r) \tag{3.25b}$$

$$V_{III}(z,r) = V_2 + \sum_{n=1}^{\infty} C_n \exp(-k_n z) J_0(k_n r). \tag{3.25c}$$

In order to find the coefficients, we must consider the conditions at the various boundaries. At the inner surfaces of the cylinders the potentials are V_1 and V_2 so, unless $A_n = C_n = 0$, the Bessel functions must all vanish for $r = D/2$. The possible values of $k_n D/2$ are therefore the roots, $j_{0,n}$, of $J_0 = 0$ and will be found in tables of Bessel functions [14]. This restriction on the values of k_n means that in region II the potential at $r = D/2$ varies linearly with z. This is not correct, but it is not a bad approximation for small values of g/D and allowances can be made. Symmetry allows us to write

$$A_n = -C_n \qquad \text{and} \qquad B_n = -B'_n$$

and from the continuity of potential at the boundary between regions I and II we have

$$A_n \exp(-k_n g/2) = B_n[\exp(k_n g/2) - \exp(-k_n g/2)]$$

which allows us to express the potentials in all three regions in terms of the single set of coefficients B_n.

$$\begin{aligned} V_I(z,r) =& V_1 + \sum_{n=1}^{\infty} B_n[\exp(k_n g) - 1]\exp(k_n z) J_0(k_n r) \\ V_{II}(z,r) =& \frac{V_1 + V_2}{2} + \left(\frac{V_2 - V_1}{g}\right) z \\ &+ \sum_{n=1}^{\infty} B_n[\exp(-k_n z) - \exp(k_n z)] J_0(k_n r) \\ V_{III}(z,r) =& V_2 - \sum_{n=1}^{\infty} B_n[\exp(k_n g) - 1]\exp(-k_n z) J_0(k_n r). \end{aligned}$$

The continuity of potential across the boundary between regions I and II implies that the radial component of the field is continuous across this boundary, but we have not used the constraint that the axial component of the field must also be continuous across this boundary,

$$\sum_{n=1}^{\infty} B_n k_n \exp(k_n g/2) J_0(k_n r) = \frac{V_2 - V_1}{2g} \qquad \text{for all } r \leqslant D/2.$$

It is not easy to apply this constraint directly, but another approach allows all the B_n to be determined.

3.6.1 The variational principle

It is a general feature of any physical system that it will adjust itself to a condition of minimum potential energy. We may therefore test the suitability of any expression for the potential by examining the energy stored in the associated field. The energy density of an electrostatic field can be written as $\frac{1}{2}\epsilon_0 E^2$. If we write the total energy of the field as the integral of the sums of the squares of the axial and radial differentials of the potentials we can determine the coefficients by equating the differentials of this integral with respect to B_n to zero. The expression for the energy, W, contains three terms, one for each of the regions, but we need only evaluate one and a half of these because of the symmetry about the central plane.

$$\begin{aligned} W = \pi\epsilon_0 \Bigg\{ & \int_{-\infty}^{-g/2} \int_0^{D/2} \left[\left(\frac{\partial V_I}{\partial z}\right)^2 + \left(\frac{\partial V_I}{\partial r}\right)^2\right] r \, \mathrm{d}r \, \mathrm{d}z \\ & + \int_{-g/2}^{0} \int_0^{D/2} \left[\left(\frac{\partial V_{II}}{\partial z}\right)^2 + \left(\frac{\partial V_{II}}{\partial r}\right)^2\right] r \, \mathrm{d}r \, \mathrm{d}z \Bigg\}. \end{aligned}$$

The evaluation of these integrals is considerably simplified by the orthogonality of the Bessel functions.

Applying the condition $\partial W/\partial B_n = 0$ gives

$$B_n = (V_2 - V_1)\frac{\exp(-k_n g/2)}{(g/2)k_n^2 D J_1(k_n D/2)}$$

where $J_1(k_n D/2)$ is the Bessel function of order one evaluated at the zeros of J_0.

The properties of the field do not depend on the scale of the system and it is convenient to express all the dimensions in terms of the diameter, D, of the cylinders.

$$G = g/D \qquad Z = z/D \qquad R = 2r/D$$

with

$$K_n = k_n D/2$$

Values of K_n and the corresponding values of $J_1(K_n)$ to ten decimal places can be found in tables [14] for $1 \leqslant n \leqslant 150$. Writing $Q_n = 1/2K_n^2 J_1(K_n)$ we have $B_n = (V_2 - V_1)Q_n \exp(-K_n G)/G$. The exponential terms involve $(\pm 2Z \pm G)$ so we write

$$a1_n = \exp[K_n(2Z + G)] \qquad \text{and} \qquad a2_n = \exp[K_n(2Z - G)]$$

and express the potentials as

$$V_I(Z, R) = V_1 + \frac{V_2 - V_1}{G}\sum_{n=1}^{\infty} Q_n J_0(K_n R)(a1_n - a2_n) \tag{3.26a}$$

$$\begin{aligned} V_{II}(Z, R) = {} & \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{G}Z \\ & + \frac{V_2 - V_1}{G}\sum_{n=1}^{\infty} Q_n J_0(K_n R)(1/a1_n - a2_n) \end{aligned} \tag{3.26b}$$

$$V_{III}(Z, R) = V_2 - \frac{V_2 - V_1}{G}\sum_{n=1}^{\infty} Q_n J_0(K_n R)(1/a2_n - 1/a1_n). \tag{3.26c}$$

Successive terms of the summations converge quite rapidly except close to the boundaries between the regions, and at axial distances greater than some $1\frac{1}{2}$ diameters only the first term makes any significant contribution. The axial potential then approaches that of the electrode as $e^{-2K_0 Z}$ where $K_0 = 2.4048$. As $G \to 0$ the exponential terms approach $2K_n G \exp(-2K_n Z)$, the $(2K_n G)$ cancels and the expressions reduce to

$$2\sum_{n=1}^{\infty} \frac{\exp(-2K_n|Z|)}{K_n J_1(K_n)} J_0(K_n R).$$

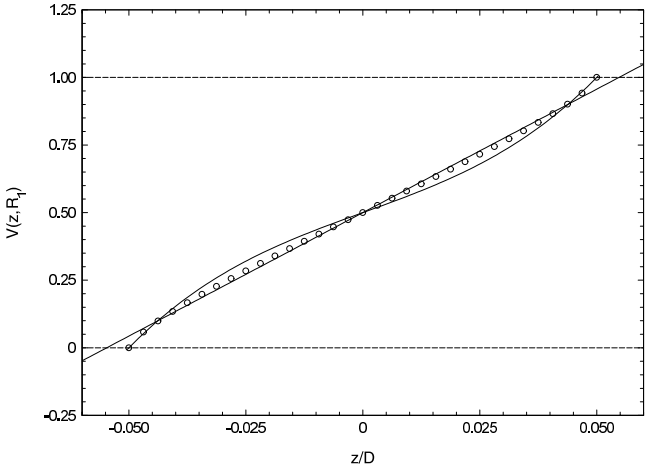

Figure 3.18 The potential in the gap between two cylinders at potentials 0 and 1. The full curve shows the 'cubic' result of Bonjour [15] and the points show the results of nine point relaxation with a lattice spacing of $D/320$. The straight line is a least squares fit to these data and its significance is discussed in the text.

3.6.2 The problem of the non-linear potential in the gap

For small gaps the departure of the true potential between the cylinders at $r = D/2$ from a linear dependence on z is not really serious, but nonetheless it is possible to make a correction in one of two ways. Instead of writing the second term of equation (3.25*b*) as $(V_2 - V_1/g)z$ a third order term can be added [15], and the variational calculation repeated with the slightly more complicated expression

$$(V_2 - V_1)\left[\left(1 - \frac{a}{2}\right)\frac{z}{g} + 2a\left(\frac{z}{g}\right)^3\right].$$

For a zero gap lens, $a = 1$, and decreases quite slowly, being about 0.75 for $g = D/2$. This is too simple to represent the true situation except, perhaps, for cylinders with extremely thin walls. Thick walled cylinders are more practicable and have the further virtues of shielding the cylindrical space from charged surfaces outside the lens system, but within the vacuum chamber.

The true variation of the potential at $r = D/2$ may be found using the finite difference method to relax the potential not only within the cylindrical surface, but outwards to a greater radius within the gap as discussed in section 3.4.2.1 in the context of aperture lenses. Figure 3.18 illustrates this variation for a gap, $g = D/10$. The cubic form is clearly a poorer representation than the simple linear variation and the true shape requires higher order terms to describe it. It

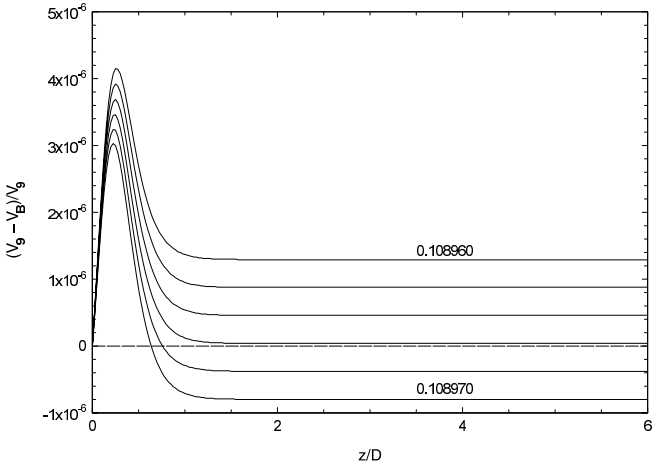

Figure 3.19 The fractional difference between the axial potential calculated using the Bessel function expansion and that calculated using nine point relaxation. Six values of the gap scaling factor, from 1.089 960 to 1.089 970 were used, the topmost curve corresponding to the smallest value.

would be quite difficult to repeat the variational calculation with fifth and seventh order terms but the straight line of slope 9.135 provides a guide to improving the approximation. Instead of using the true value of the gap in the expressions for the potential, we can multiply the this by a scaling factor slightly larger than one such that the axial potential agrees with that calculated by the nine point relaxation method to a high degree of accuracy except close to the actual gap.

We show in figure 3.19 the fractional difference between the potential calculated using a range of scaling factors, and that calculated by the nine point relaxation method for a lens represented by arrays 160×3200 for the cylinder and 160×160 for the gap between the cylinders. Interpolation between these values of gap scaling factor suggests that a value of 1.089 662 will give quite excellent agreement for $z/D > 1$ and very good agreement close to the gap. The greatest fractional difference is 3.4×10^{-6} at $D/4$.

In addition to providing a value for the gap scaling factor for a nominal gap of $D/10$, figure 3.19 shows that the nine point method gives an accurate description of the axial potential. The agreement extends over the whole volume of the lens cylinder as is shown in figure 3.20. Contours of constant fractional difference show that, out to a radius of $D/4$, the agreement is better than 10^{-5}. The disagreement for smaller values of z/D is due to the approximation used in the Bessel expansion calculation.

3.6.3 Calculation of the potential

A Pascal function which calculates the potential at the point (Z, R) using

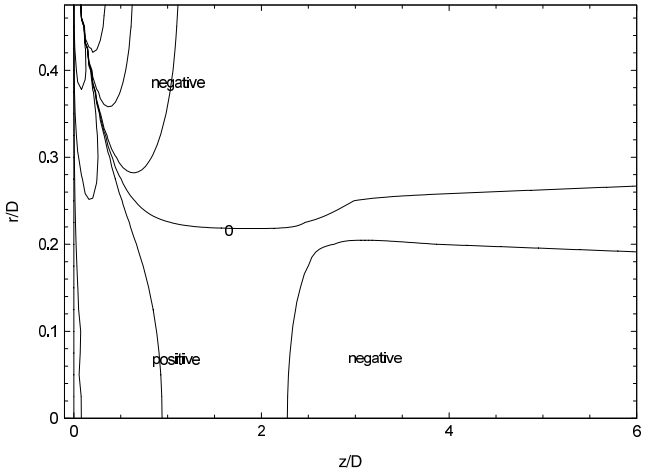

Figure 3.20 The fractional difference between the potential calculated using the Bessel function expansion and that calculated using nine point relaxation with a gap scaling factor of 1.089 662. The space is divided by the contours of zero value into region in which the value of the fractional difference is positive or negative. The other contours correspond to values of $\pm 10^{-6}$, 10^{-5} and 10^{-4}.

equations (3.26) is shown on the following page. It uses a further function **BessJ0(x)** which can be taken directly from an example in *Numerical Recipes* [7]. The values are appropriate for cylinder potentials of zero and unity and for a *nominal* gap of $D/10$. The main program must supply values for $K[n]$, the nth zero of the Bessel function of order zero, and $Q[n]$. Both of these can be determined from values in tables [14].

At each position the potential, V, is calculated by summing terms $\mathrm{d}V[c]$ until the absolute value of a term is less than some limiting value (taken here as 10^{-8}) or the number of terms exceeds 150, when a rounding-off procedure is applied to estimate the sum to infinity.

The series converges rather slowly close to the boundaries between the regions. If it has not converged after 150 terms, we need to invoke the round-off procedure. By considering the continuity of the gradient at the boundary between regions I and II, it is easy to show that

$$\sum_{n=1}^{\infty} \frac{1}{K_n J_n} \equiv \sum_{n=1}^{\infty} \frac{1}{k[n]j[n]} = 0.5$$

but the sum to 150 terms is, at $0.471\,15\cdots$, significantly different from this. The sums to odd and even numbers of terms differ from the true sum by amounts which alternate in sign and decrease progressively. A simple average introduces some

```
FUNCTION BesselPot(R,Z:double):double;
CONST
  Gap=0.109362;
VAR
   Limit,Z1,Z2,V0:double;
   C:integer;
   S:ARRAY[0..156] of double;
   dV:ARRAY[0..150] of double;
PROCEDURE RoundOff(C:integer);
BEGIN
     S[C+3]:=((S[C]*S[C-2]-SQR(S[C-1]))/(S[C]-2*S[C-1]
           +S[C-2]));
END;
BEGIN
     Limit:=0.00000001;
     Z1:=2*Z + Gap;     Z2:=2*Z - Gap;
     C:=1;        S[0]:=0;     dV[0]:=1;
     WHILE ABS(dV[C-1]) > LIMIT DO
     BEGIN
IF Z > Gap/2 THEN BEGIN
       dV[C]:=Q[C]*BessJ0(K[C]*R)*(exp(-K[C]*Z1)
             -exp(-K[C]*Z2))/Gap;
               V0:=1;  END
ELSE IF Z < -Gap/2 THEN BEGIN
       dV[C]:=Q[C]*BessJ0(K[C]*R)*(exp(K[C]*Z1)
             -exp(K[C]*Z2))/Gap;
               V0:=0;  END
ELSE BEGIN
       dV[C]:=Q[C]*BessJ0(K[C]*Z)*(exp(-K[C]*Z1)
             -exp(K[C]*Z2))/Gap;
               V0:=0.5+Z/Gap;  END;
       S[C]:=S[C-1]+dV[C];     V:=V0+S[C];
       IF C < 150 THEN C:=C+1
                  ELSE BEGIN
                          RoundOff(148);
                          RoundOff(149);
                          RoundOff(150);
                          RoundOff(153);
                          V:=V0+S[156];
                          dV[149]:=0;
                        END;
     END;
     BesselPot:=V;
END;
```

bias as a result of this progressive decrease, but a mean of three successive terms is really very good. The round-off procedure used in this program is much more powerful than is really necessary and could probably be applied safely to estimate the sum to infinity after only the first 30 or so terms had been calculated. It is used to extrapolate from the sums to 146 to 150 terms by using

$$S[C+3] = \frac{S[C] \times S[C-2] - S^2[C-1]}{S[C] - 2S[C-1] + S[C-2]}$$

with $C = 148,\ 149$ and 150 to generate pseudosums, $S[151],\ S[152]$ and $S[153]$ which are then averaged again to give a final pseudosum, $S[156]$, which is within 10^{-10} of 0.50 for the exact boundary case. Away from the boundaries, the series converge quite quickly: 20 terms being enough by $|z/D| = 0.15$.

The Bessel function expansion is by far the fastest method for calculating the potential in a two cylinder lens. On a 450 MHz Pentium III machine, the potential can be evaluated for a lattice of 160 radial and 3200 axial points in a quarter of a minute. The nine point finite difference method requires about half an hour!

4 The optics of simple lens systems

4.1 Aperture lenses

One might almost say that aperture lenses are the 'traditional' particle lens. A great variety of detailed shapes and proportions have been used which makes it difficult to prepare tables of focal data for general application. In the context of low energy beams there is no need for smoothly rounded electrodes: simple thin discs with circular holes can be used and, provided that there is some agreement about the spacing of the discs, it should be possible to make systems of two or three apertures which behave in a predictable fashion. Figure 3.10 shows the important dimensions of such systems. For the purposes of scaling, it is usual to express all the dimensions in terms of the diameter, D_1, of the apertures themselves. The significance of the outer diameter of the discs, D_3, was discussed earlier in the context of establishing a simple boundary condition to use in the relaxation calculation, but it has a further, practical, significance in that the lens system does not exist in a vacuum, but rather in a vacuum *system* which has walls which will either be of metal and therefore have some definite potential, or of an insulating material and have a rather indefinite potential. In both cases it is important that this should not influence the potential distribution within the lens and it is as well to ensure that $(D_3 - D_1)/A > 3$.

A similar argument requires that the first and last apertures be mounted on cylinders to guarantee uniform potentials in the object and image spaces, and the diameter, D_2, of these cylinders will affect the potential distribution and, therefore, to some extent the focusing properties. Figure 4.1 shows the axial potential distribution in a two aperture lens with $A = D_1/2$ and a number of values of D_2/D_1. Near the centre of the lens the distributions are very similar, but with increasing distance from the centre the potential tends towards the simple exponential decay characteristic of the dominant term in the Bessel function expansion, $V(z, 0) \propto \exp(-k_0 z/\frac{1}{2}D_2)$. Most of the calculations were made with the length of the D_2-cylinder great enough that the potential fell smoothly, but one example is included of a cylinder of length $4D_1$, and therefore only $4D_2/3$, to illustrate the way in which the end-plate can influence the potential distribution.

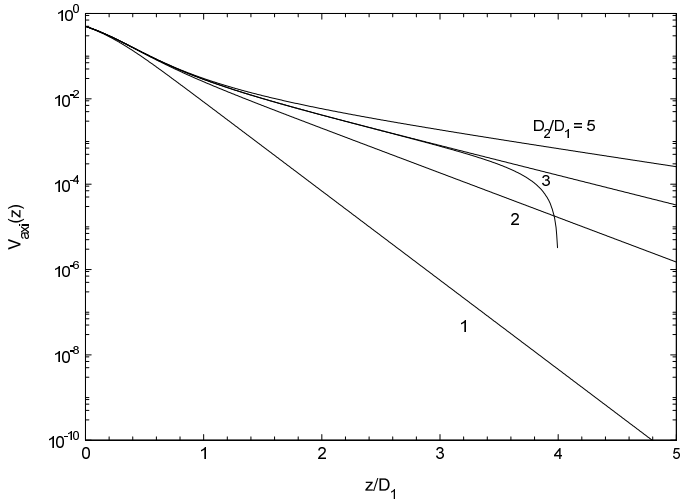

Figure 4.1 Axial potential distributions for two aperture lenses having $A = D_1/2$. Values of the ratio D_2/D_1 are shown on the diagram. The potentials all have the value 0.5 at the lens centre, but fall with increasing axial distance, z, as $\exp(-4.8096(z/D_2))$. The potentials approach these asymptotes from above except in the case of the cylinder lens. Is there a value of D_2/D_1 for which the potential falls exponentially for all z?

The different potential distributions away from the apertures themselves affect the properties of the lens as illustrated in figure 4.2. This shows the first focal length, f_1, mid-focal distance, F_1, and spherical aberration coefficient, C_{s0}, for a voltage ratio, $V_2/V_1 = 10$, as functions of the ratio D_2/D_1, and includes the case of the cylinder lens of the same spacing, which one might consider to be a rather extreme aperture lens having $D_2/D_1 = 1$. The overall effect on the focal properties is not large, but the lower aberration of the cylinder lens is quite noticeable. Changes of a similar scale result if the separation of the apertures is doubled, but whereas this is a fairly obvious effect, the effect of changes in D_2 is a little more subtle. The thickness of the aperture discs also plays a part, but not on the same scale. The data presented in the **LENSYS** program are for $D_2/D_1 = 3$ which represents a compromise between reduced dependence at large values of the ratio and sheer physical size of the electrodes.

4.1.1 Three-aperture lenses

As is the case with all lenses with three elements, the principal function is the decoupling of the focal properties of the lens, in particular the behaviour between conjugates, from the ratio of the potentials of the image and object spaces. Data for this sort of application are generally presented in the form of plots of the ratio V_2/V_1 against V_3/V_1. These plots are usually called 'focal loci' or 'zoom

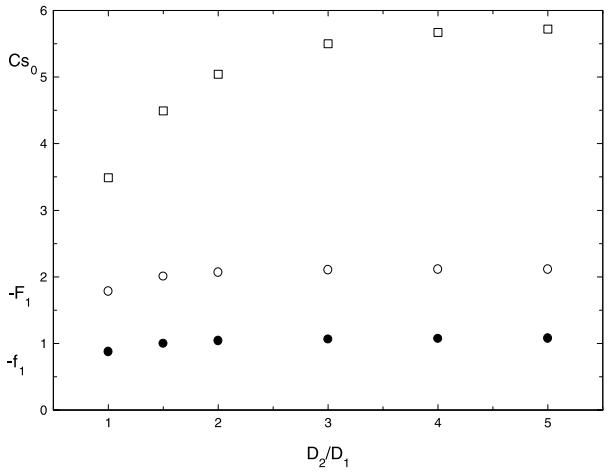

Figure 4.2 Focal lengths (○), mid-focal distances (●) and spherical aberration coefficients (□) for two aperture lenses having $A = D_1/2$ and $V_2/V_1 = 10$.

curves' and lenses used in this fashion are referred to as 'zoom lenses', but the behaviour is more restricted than the use of this term in light optics would imply. We shall return to this topic in section 4.6.1.

4.1.1.1 The Einzel lens. Nowadays this term is understood to refer to a three element lens operated symmetrically, with $V_3 = V_1$ and with V_2 taking any appropriate value. Historically the Einzel, or unipotential, lens signified a system with V_2 held at the potential of the cathode. The major advantage of this is that the focal properties, depending as they do on voltage *ratios,* are the same for all values of $V_1 = V_3$. The disadvantage is that the aberrations of this lens are greater than those of the symmetric lens with $V_2 > V_{1,3}$.

The potential distribution in an Einzel lens is different in character from that shown for the two cylinder lens in figure 2.1(b) and we illustrate this in figure 4.3. Apart from the bowing out of the equipotentials which occurs for any aperture lens, the potential close to the centre forms a saddle where the potential rises (say) in the r-direction and falls in the $\pm z$-direction. This general form will apply to any three aperture lens for which V_2 is outside the range spanned by V_1 and V_3. One of the equipotentials crosses the axis at the centre and it is simple to show that the angle between this equipotential (the value of which will depend on the thickness of the aperture plates) and the axis is $\pm 54.7°$. Expanding the axial potential as

$$V(z, 0) = V_0(z, 0) + zV_0'(z, 0) + \tfrac{1}{2}z^2 V_0''(z, 0) + \cdots$$

and substituting this into equation (2.2) gives

$$V(z, r) = V_0(z, 0) + zV_0'(z, 0) + \tfrac{1}{2}z^2 V_0''(z, 0) - \tfrac{1}{4}r^2 V_0''(z, 0) + \cdots.$$

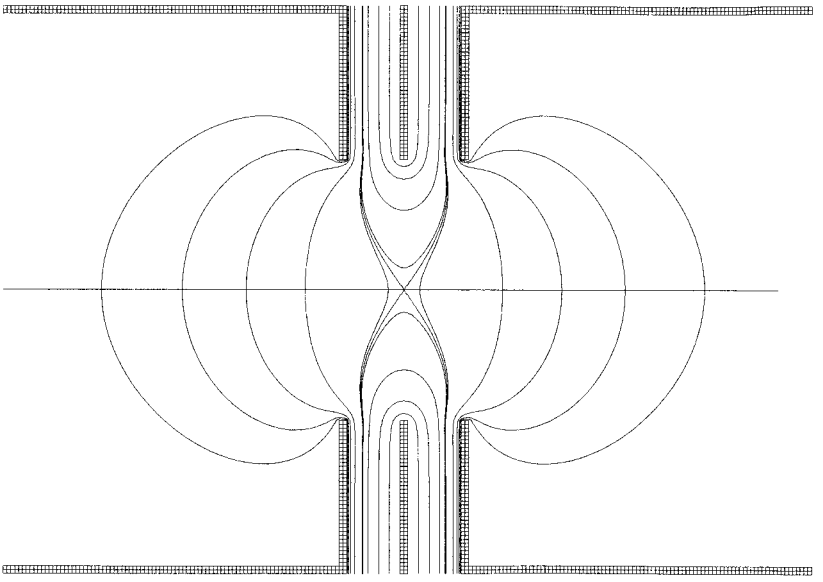

Figure 4.3 Equipotentials in an Einzel lens. Near the centre of the lens the equipotentials are at 0.300, 0.294 and 0.288 of the potential difference between the outer and centre electrodes. Close to the axis the equipotentials are hyperbolæ, and at the centre they form a pair of straight lines which cross the axis at angles of $\pm 54.7^\circ$.

$V_0'(z, 0) = 0$ at the saddle point and $V(z, r) = V_0(z, 0)$ along the equipotential, so the equation of the equipotential close to the saddle point is $\frac{1}{2}z^2 = \frac{1}{4}r^2$ from which the result for the angle follows immediately.

4.2 Cylinder lenses

Cylinder lenses offer a number of advantages over aperture lenses of the same inner diameter. For a given voltage ratio they are a little stronger and have smaller aberration coefficients, but there are two quite important mechanical advantages as well. They are easy to make and to mount: it is easier to turn cylinders on a lathe than to punch clean-edged holes in metal sheet. It is also simple to make lens systems with rather small inter-electrode gaps. This is an important factor in keeping the potential distribution in the gap at the inner surface of the electrodes close to that assumed in the calculation of the entire potential distribution and also ensures good screening from any stray electric fields around the lens structure.

The potential and field distributions in a system of several coaxial cylinders can be found quite adequately by the superposition of those for contiguous pairs, provided that the gaps between the cylinders are small (say, $\not> 0.2\ D$) and the cylinders themselves are rather longer than the gaps. In practice it is common

to use cylindrical elements no shorter than $D/2$ (gap centre to gap centre) with gaps of $D/10$. This allows lens systems having a large number of cylinders to be designed and analysed, procedures which would be much more difficult with aperture electrodes. If the gaps were large, it would be necessary to solve each geometry with appropriate symmetric and antisymmetric potentials on all electrodes, as in the case of the three aperture lens.

4.3 The calculation of ray paths

There are no purely analytic solutions to this problem, only various procedures for the numerical integration of the path starting from a known position and slope at some point. The choice of method depends on the extent and quality of the data. If only the axial potential is known, then integration of the Picht equation can be done, and there is one procedure which is particularly well suited to this. On the other hand, if the potential and its derivatives are known quite accurately throughout the space, a direct application of the laws of simple dynamics will give good results. If the potential is known everywhere, but to limited accuracy, various standard procedures are available. However, care should be taken to distinguish between the precision of the integration procedure and the accuracy of the trajectory. Ray tracing in the *correct* potential will give results which show the effects of aberrations, while integration of the Picht equation will give only the paraxial, Gaussian, behaviour though corrections can be applied and aberration coefficients calculated even in this case.

4.3.1 Ray tracing in a known potential

If the potential is not known really well throughout the space, the best approach is to use some standard expression for the integration of the differential equations, but if it and its derivatives are well known an analysis by direct application of the laws of dynamics using expansions of the coordinates and velocities about each point will give good results [16].

4.3.1.1 Integration of the equation of motion. Taking the time, t, to be the independent variable we write the coordinates as

$$z(t) = z_0 + t\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)_0 + \frac{t^2}{2}\left(\frac{\mathrm{d}^2 z}{\mathrm{d}t^2}\right)_0 + \frac{t^3}{6}\left(\frac{\mathrm{d}^3 z}{\mathrm{d}t^3}\right)_0 + \cdots \tag{4.1}$$

with a similar expression for $r(t)$, where $z(t)$ and $r(t)$ are the axial and radial positions of a particle, having initial coordinates denoted by the '0' subscripts, after a short time interval t. The components of the velocity follow from differentiation of these equations

$$\frac{\mathrm{d}z(t)}{\mathrm{d}t} = \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)_0 + t\left(\frac{\mathrm{d}^2 z}{\mathrm{d}t^2}\right)_0 + \frac{t^2}{2}\left(\frac{\mathrm{d}^3 z}{\mathrm{d}t^3}\right)_0. \tag{4.2}$$

The accelerations are given by

$$\left(\frac{\mathrm{d}^2 z}{\mathrm{d}t^2}\right)_0 = \frac{e}{m}\frac{\partial V(z,r)}{\partial z}$$

and the third time-derivatives of the coordinates are therefore

$$\left(\frac{\mathrm{d}^3 z}{\mathrm{d}t^3}\right)_0 = \frac{e}{m}\left(\frac{\partial^2 V}{\partial z^2}\right)_0\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)_0 + \frac{e}{m}\left(\frac{\partial^2 V}{\partial z \partial r}\right)_0\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)_0.$$

Substituting these expressions into equations (4.1) and (4.2), and introducing a 'reduced time' $\tau = t\sqrt{2e/m}$ to avoid the '$\frac{e}{m}$' which occurs in nearly every term, we construct the following recurrence relationships.

$$\begin{aligned} z(t) &= z_0 + \tau\left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)_0 + \frac{\tau^2}{4}\left(\frac{\partial V}{\partial z}\right)_0 \\ &\quad + \frac{\tau^3}{12}\left[\left(\frac{\partial^2 V}{\partial z^2}\right)_0\left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)_0 + \left(\frac{\partial^2 V}{\partial z \partial r}\right)_0\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)_0\right] \\ \frac{\mathrm{d}z(t)}{\mathrm{d}\tau} &= \left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)_0 + \frac{\tau}{2}\left(\frac{\partial V}{\partial z}\right)_0 \\ &\quad + \frac{\tau^2}{4}\left[\left(\frac{\partial^2 V}{\partial z^2}\right)_0\left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)_0 + \left(\frac{\partial^2 V}{\partial z \partial r}\right)_0\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)_0\right] \end{aligned}$$

with two similar expressions for the radial coordinate and velocity.

Only expressions for the potential which can be evaluated and differentiated analytically at any point, such as those expressed in terms of Bessel functions or elliptic integrals, are likely to be good enough to yield accurate values for the second derivative required for this type of analysis. The calculations are surprisingly simple and we illustrate this using the potential in the gap region of a two cylinder lens developed in section 3.3.6 (equation (3.26*b*))

$$V_{II}(Z,R) = \frac{V_1+V_2}{2} + \frac{V_2-V_1}{G}Z + \frac{V_2-V_1}{G}\sum_{n=1}^{\infty} Q_n J_0(K_n R)(1/a1_n - a2_n).$$

The five differential expressions can be written as

$$\begin{aligned} \frac{\partial V}{\partial Z} &= \frac{(V_2-V_1)}{G} - \frac{2(V_2-V_1)}{G}\sum_{n=1}^{\infty} Q_n K_n J_0(K_n R)(1/a1_n + a2_n) \\ \frac{\partial V}{\partial R} &= -\frac{2(V_2-V_1)}{G}\sum_{n=1}^{\infty} Q_n K_n J_1(K_n R)(1/a1_n - a2_n) \\ \frac{\partial^2 V}{\partial Z^2} &= \frac{4(V_2-V_1)}{G}\sum_{n=1}^{\infty} Q_n K_n^2 J_0(K_n R)(1/a1_n - a2_n) \\ \frac{\partial^2 V}{\partial R^2} &= -\frac{2}{R}\frac{\partial V}{\partial R} - \frac{\partial^2 V}{\partial Z^2} \\ \frac{\partial^2 V}{\partial Z \partial R} &= \frac{4(V_2-V_1)}{G}\sum_{n=1}^{\infty} Q_n K_n^2 J_1(K_n R)(1/a1_n - a2_n). \end{aligned}$$

Very similar expressions apply in the outer regions, the principal differences arising from the different exponents.

Simple modifications of the BesselPot function of section 3.6.3 may be used to calculate these differentials. A Pascal procedure for the calculation of the J_1 Bessel function required for the R-derivatives is given in [7].

The accuracy of the ray integration depends on the interval in reduced time, $\Delta\tau$, which is related to the step length, Δl, by $\Delta\tau = \Delta l/V^{1/2}$ and it is important to choose a suitable value. Too short an interval will waste computer time, but the optimum interval will depend on the strength of the field and so it needs to be adjusted dynamically. The accuracy of the calculations is open to a simple test: the difference in potential between the initial and final points of each step should be equal to the difference in the squares of the velocities of the particle multiplied by $m/2e$. If the fractional difference found after a step is greater than some chosen value, then the step should be recalculated with $\Delta\tau$ halved. If the difference exceeds some smaller criterion, $\Delta\tau$ may be doubled for the next step.

4.3.1.2 Runge–Kutta methods. Unless we have a good enough potential we cannot make proper use of terms beyond the first order of the Taylor expansion of equation (4.1) and need to find some other way to make allowance for the higher order terms. The Runge–Kutta methods are based on the replacement of these terms by first order expansions at nearby points. Figure 4.4 illustrates the steps required to apply these methods. The function describing some aspect of the trajectory is represented by a curve starting at the point A, (x_n, y_n), and we wish to find the y-coordinate at C, for which $x_{n+h} = x_n + h$. We can write this as $y_{n+1} = y_n + k$ where k is h times some mean gradient. The very first approximation would be to write $k = k_1 = hy'(x_n, y_n)$ where the gradient of the function is evaluated at the point $A(x_n, y_n)$. This would take us to the point D_1 and is plainly not a good approximation.

However, common sense and the mean value theorem tell us that *at some point* along the curve the tangent is parallel to the chord AC joining the initial and the true final points. The next stage of approximation is therefore to evaluate k at the mid-point, B_1, of the chord AD_1 and write $k = k_2 = hy'(x_n + \frac{h}{2}, y_n + \frac{k_1}{2})$, which will take us to the point D_2. This is a significantly better approximation and the errors are of order h^3: it is therefore referred to as the second order Runge–Kutta method. It is worth making two further steps of this sort, because the gain in precision outweighs the complexity of the calculation, but beyond that point one should really question whether the potential function is good enough to justify the effort. The four successive values for k can be summarized as

$$
\begin{aligned}
k_1 &= hy'(x_n, y_n) && \text{at A}\\
k_2 &= hy'(x_n + h/2, y_n + k_1/2) && \text{at B}_1\\
k_3 &= hy'(x_n + h/2, y_n + k_2/2) && \text{at B}_2\\
k_4 &= hy'(x_n + h, y_n + k_3) && \text{at D}_3.
\end{aligned}
$$

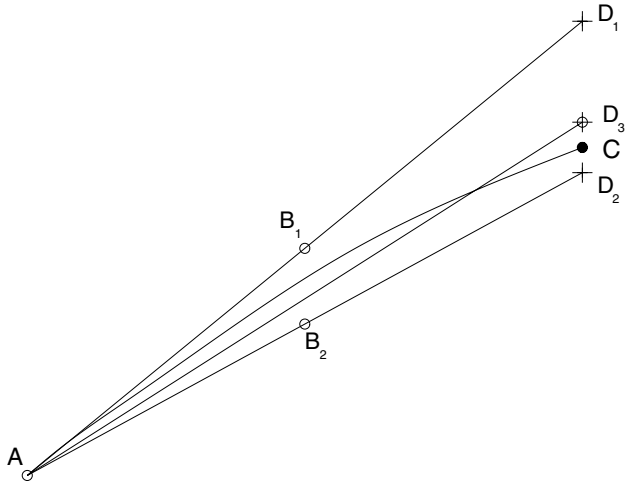

Figure 4.4 A schematic illustration of the steps in the fourth order Runge–Kutta method. AC represents the real trajectory and AD_1 is the tangent to the trajectory at A. The gradient of the function at B_1 is used to find D_2, the second approximation to C, then the gradients at B_2 and D_3 are evaluated and a weighted sum of the four gradients is then a good approximation to that of the chord AC.

These expressions have a nice symmetry and we write

$$k = \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6}$$

where the weights given to the k_i have been chosen such that the third and fourth order terms in the Taylor expansions cancel and the errors are now of order h^5 giving the fourth order Runge–Kutta method.

To apply this to the tracing of ray paths, we write the axial, z_n, and radial, r_n, positions and the corresponding velocities, u_n and v_n respectively, in terms of the reduced time, τ, as before (i.e. $u_n = \frac{dz}{d\tau}$ etc) and apply the recurrence relations to all four variables.

$$z_{n+1} = z_n + j \qquad r_{n+1} = r_n + k \qquad u_{n+1} = u_n + l \qquad v_{n+1} = v_n + m$$

where j, k, l and m are the appropriate weighted sums. The individual approximations are

$$\begin{aligned}
j_1 &= hu_n & l_1 &= -hE_z(z_n, r_n)/2 \\
j_2 &= h\left(u_n + \frac{l_1}{2}\right) & l_2 &= -hE_z\left(z_n + \frac{l_1}{2}, r_n + \frac{j_1}{2}\right)/2 \\
j_3 &= h\left(u_n + \frac{l_2}{2}\right) & l_3 &= -hE_z\left(z_n + \frac{l_2}{2}, r_n + \frac{j_2}{2}\right)/2 \\
j_4 &= h(u_n + l_3) & l_4 &= -hE_z\left(z_n + l_3, r_n + j_3\right)/2
\end{aligned}$$

with a similar set for the radial motion. h is an interval of reduced time, and E_z and E_r are the axial and radial components of the electric field strength which must be found at all the indicated points by numerical differentiation of the potential.

4.3.2 *Integration of the Picht equation*

The Picht equation describes the evolution of a reduced radius defined by $R(z) = r(z)V^{1/4}(z, 0)$ as a function of z. It involves the axial potential, $V(z, 0)$ and its first derivative and contains only the second derivative of $R(z)$. There is a standard procedure for the integration of such equations which is faster than fourth order Runge–Kutta with the same step size and is effectively of higher order. We write the Picht equation as

$$\frac{\mathrm{d}^2 R(z)}{\mathrm{d}z^2} = -T^*(z)R(z)$$

where

$$T^*(z) = \frac{3}{16}\left(\frac{V'(z, 0)}{V(z, 0)}\right)^2.$$

$T^*(z) = \frac{3}{16}T(z)^2$ where $T(z)$ was introduced in chapter 2. Some authors use $T(z)$ where we use $T^*(z)$, but it is necessary to distinguish the two symbols because $T(z)$ itself will appear in expressions for the aberration coefficients. For compactness we shall omit the explicit references to z in R and T^* in what follows.

We shall integrate the equation using equal steps, h, in z and so write for the values of R at adjacent steps

$$R_{n\pm1} = R_n \pm hR'_n + \frac{h^2}{2}R''_n \pm \frac{h^3}{6}R'''_n + \frac{h^4}{24}R^{(4)}_n \pm \cdots$$

and rearrange this to obtain

$$\frac{R_{n+1} - 2R_n + R_{n-1}}{h^2} = R''_n + \frac{h^2}{12}R^{(4)}_n + \cdots.$$

$R''_n = -(T^*R)_n$ and the fourth differential of R_n can be written in terms of the second differential of $(T^*R)_n$ as

$$\begin{aligned} R^{(4)}_n &= \frac{\mathrm{d}^2(-T^*R)_n}{\mathrm{d}z^2} \\ &= -\frac{(T^*R)_{n+1} - 2(T^*R)_n + (T^*R)_{n-1}}{h^2} \end{aligned}$$

using an exactly similar expansion of T^*R, but neglecting the fourth order term in the expansion, giving

$$\begin{aligned} \frac{R_{n+1} - 2R_n + R_{n-1}}{h^2} &= -(T^*R)_n \\ &\quad - \frac{h^2}{12}\frac{[(T^*R)_{n+1} - 2(T^*R)_n + (T^*R)_{n-1}]}{h^2} \end{aligned}$$

and, rearranging this equation, we finally have

$$R_{n+1} = \frac{2R_n - R_{n-1} - \frac{h^2}{12}[(T^*R)_{n-1} + 10(T^*R)_n]}{[1 + \frac{h^2}{12}T^*_{n+1}]} \tag{4.3}$$

This equation is known as Numerov's algorithm† and it is very effective within its limitations. Two values of $R(z)$ are required to start the process, but this is not a problem if the starting point is in a field free region. If this is not the case, some other method has to be used to start the process, and a Runge–Kutta method would normally be used. The step size has to be chosen carefully as there is no easy way to adjust it dynamically, since the particle velocity is not calculated at any stage. This means that a step size suitable for the worst part of the trajectory has to be used throughout.

4.3.2.1 The Runge–Kutta method applied to the Picht equation. As usual, the Runge–Kutta method requires parameters to be calculated in the middle of the integration steps. If the values of T^* are known at intervals $(h/2)$ the Runge–Kutta method must use a step size of h while the Numerov method can use the smaller step giving it a 16-fold advantage in precision. The expressions for the reduced radius R and its gradient R' are

$$\begin{aligned} R_{n+1} &= R_n - R_n\frac{h^2}{6}\left(T^*_n + 2T^*_{n+1/2} - \frac{h^2}{4}T^*_nT^*_{n+1/2}\right) \\ &\quad + R'_n\left(1 - \frac{h^2}{6}T^*_{n+1/2}\right) \\ R'_{n+1} &= -R_n\frac{h}{6}\left(T^*_n + 4T^*_{n+1/2} - \frac{h^2}{2}T^*_nT^*_{n+1/2}\right) \\ &\quad + R'_n\left(1 - \frac{h^2}{3}T^*_{n+1/2}\right) - \frac{h}{6}T^*_{n+1}R_{n+1}. \end{aligned}$$

Notice that it is necessary to calculate R_{n+1} before R'_{n+1}.

4.4 Real and asymptotic cardinal points

Particle lenses differ from those for light in having no abrupt boundaries between regions of different refractive index, so there is no clear cut point at which the lens begins or ends. Though the ray paths are asymptotically straight, away from the lens, a particle emerging from the lens field may cross the axis before this asymptotic region is reached. If the particle had approached the lens parallel to the axis, this crossing point will be a focal point and the mid-focal distance will

† It is also known as the Manning–Millman and the Fox–Goodwin methods.

be well defined, but how should the focal length be defined? In chapter 1 the focal length was defined as the distance of the focal point from a principal plane which was itself defined by the intersection of asymptotes. This is not appropriate to the present case and we must define a *real* principal plane in terms of the incident asymptote and a tangent to the ray path at the focal point. These real cardinal points are of limited use in the analysis of lens behaviour because for an object not at infinity the image will be formed beyond the focal point and therefore closer to the asymptotic region, if not fully in it. Fortunately, the distinction is necessary only for lenses which are stronger than are commonly used except, perhaps, for microscope objectives.

4.4.1 The determination of the cardinal points by ray tracing

If the path of a ray starting some three to four diameters from the lens is traced, using any of the methods outlined in section 3 the position of the focal point can be found by noting the values of z for which the radius, r, or reduced radius, R, changes sign and interpolating to find the position of the crossing point. This will be the true focal point, but, unless the lens is very strong indeed, it will be very close to the asymptotic focal point. If the ray path has not crossed the axis by the end of the integration, the focal point can be located by extrapolating the last few points calculated. The corresponding principal point is found from the slope of either the tangent to the path at the focal point or of the asymptote. With a good potential the effects of spherical aberration can be seen by changing the off-axis distance of the incident ray, and the spherical aberration coefficient can be determined. If the trace is done by integrating the Picht equation, it is not necessary to convert back to real radius as both r and R are zero at the crossing point and the asymptote is only reached in a region of constant potential. Nonetheless, it is interesting and instructive to make the conversion as the approach to the asymptotes has a rather different appearance.

4.5 Windows and pupils

While it is the paths of individual rays which are calculated, measurements are made with *beams* of particles and these have finite lateral and angular extent. These beams must be defined by suitable apertures, and in figure 4.5 we illustrate some of the terms and parameters involved. The beam is defined by two apertures, called conventionally the window and the pupil. The window is regarded as the object on which the lens acts to produce an image and the distinction is sometimes made between the object window and the image window. The pupil, or angle stop, defines the range of angles which may enter the lens from any point in the window. Naturally the lens produces an image of the pupil also, and the position and size of the image pupil are important parameters of an optical system. In figure 4.5 the lens is represented as a thin lens with coincident principal planes purely to avoid complexity in the diagram. The lateral dimensions are grossly exaggerated. The

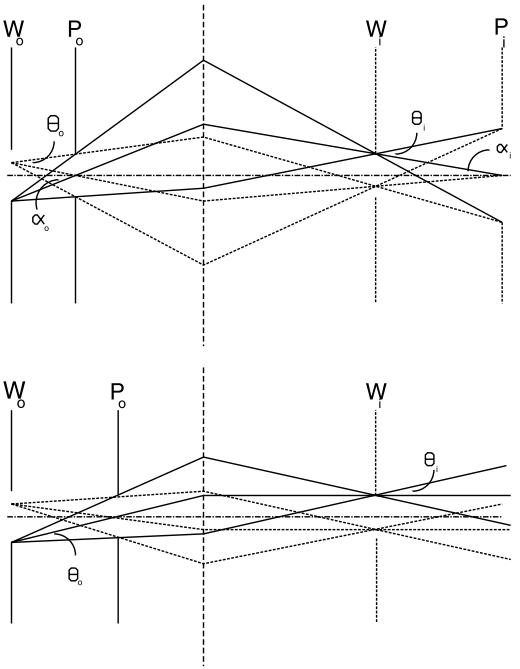

Figure 4.5 Diagram to illustrate the formation of the image of an entrance, or object, window W_0, by a bundle of rays limited by an entrance pupil, P_0. The semi-angular divergence of the bundle of rays is specified by the *pencil angle,* θ_0, and the maximum inclination of the central ray of the bundle to the axis in the plane of the entrance pupil is the *beam angle,* α_0. The lens forms images of the object-side window and pupil, and the lower diagram shows the special case where the object-side pupil is at the first focal point, giving an image pupil at infinity and zero beam angle on the image side.

upper part of the diagram shows the paths of a number of rays from the bottom of the object window and a point within it which pass through the object pupil at its upper and lower edges and also through the centre of the pupil. In the image space the rays from each point in the object window cross at the corresponding point in the image window and those rays which passed through particular points of the object pupil pass through the corresponding parts of the image pupil. Two sets of angles are indicated: α and θ. α is known as the beam angle and is the symbol used for the angle between the principal rays in the construction of figure 1.3. It is the angle subtended at the centre of the pupil by the radius of the window and is therefore the angle at which the central ray from the edge of the window crosses the axis in the plane of the pupil. θ is the pencil angle, describing the angular spread of a pencil of rays from a point in the window, and is the angle subtended at the centre of the window by the radius of the pupil. Suffices distinguish the object

and image space values of these angles. If the radii of the window and pupil are written as r_w and r_p respectively and the distance between them is l then the angles can be expressed as

$$\alpha = \frac{r_w}{l} \qquad \theta = \frac{r_p}{l}.$$

The lower part of figure 4.5 shows a special case. The object pupil is at the first focal point of the lens and is therefore imaged at infinity. The image side beam angle is now zero and the total angular divergence of the beam, which can be written as $(|\alpha| + |\theta|)$, is a minimum.

4.5.1 The energy-add lens

If a real image of a window is formed at the first principal plane of a lens, the emergent rays appear to come from a virtual image of the same size at the second principal plane. These rays may be focused to give a final, real image. The simplest practical system consists of four cylinders. The central two cylinders form the energy-add lens proper, and exploit the fact that the positions of the principal planes of a two cylinder lens do not change much over a fairly wide range of voltage ratios. The lens formed by the first two cylinders produces the image at the first principal plane of this central lens, and that formed by the last two cylinders gives the final image. The voltage ratios V_2/V_1 and V_4/V_3 are held constant, but the ratio V_3/V_2 is changed, *adding* energy $e(V_3 - V_2)$ to the particle. This apparently simple method of maintaining the position of the final image over a range of overall voltage ratio, V_4/V_1, has major drawbacks for all but the crudest applications. While the imaging of the entrance window may be quite well controlled, there is no control over the position and size of the pupil image formed by the first lens. The action of the second and third lenses will only make matters worse, most probably leading to excessive beam angles at the final image. The presence of an intermediate image is also likely to cause increased aberrations. The simple multi-element lenses described in the next section are much better behaved.

4.6 Multi-element lenses

4.6.1 Zoom lenses

A lens having two elements allows a conjugate focus condition to be met in just two ways, with an accelerating or a decelerating voltage ratio. The addition of a third element allows the condition to be met over a wide range of overall voltage ratios by a suitable choice of the potential of the third, centre, electrode. The focal locus for one such lens is shown in figure 4.6. The magnification of the lens can only be controlled in the most general fashion by, for example, selecting the lengths of the cylinders to give particular values at a limited number (three, at the most) of points on the locus. In order to make a true zoom lens, one in which the magnification for a given pair of conjugates can be controlled independently of the overall voltage

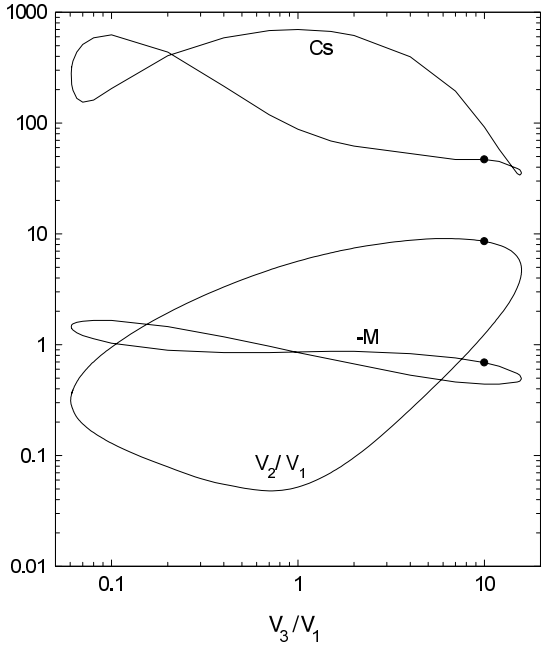

Figure 4.6 Focal locus for a three cylinder lens having elements of length 3.0, 0.75 and 2.5 diameters. The variation of the magnification and of the spherical aberration round the locus are also shown. A ● identifies corresponding points on the three loci.

ratio, a fourth element is necessary. It is difficult to obtain much flexibility with aperture lenses because the electrodes have to be comparatively close together but, with cylindrical electrodes, the distance between the two outer electrodes, which determine the overall voltage ratio, can be usefully large. The two central electrodes must have potentials which combine in some way to satisfy the focus condition, but this can be done in many ways and one might consider the two potentials to have some average magnitude (though not a simple arithmetic mean) and also a 'centre of action' which can be moved to the left or right by changing their *ratio*. Figure 4.7 illustrates this for two four cylinder lenses with symmetric conjugates separated by six diameters and with their central two elements of length D in one case and $D/2$ in the other. This figure shows the magnifications which may be obtained for an overall voltage ratio of unity. The lens with the longer central elements offers a range of magnifications from 0.67 to 1.5 while the other will only cover from 0.87 to 1.15. Notice that the extreme magnifications are obtained with both centre electrodes at higher potentials than the outer ones and not, as one might have expected, with one of the centre elements at the outer potential. It is left to the reader to investigate (using the **LENSYS** program) other

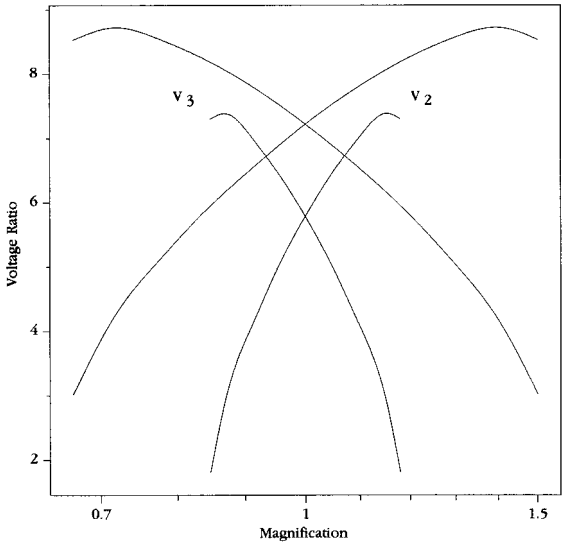

Figure 4.7 The potentials of the two centre elements of four cylinder lenses of total length six diameters required to give a range of magnifications. The upper curves relate to a lens with centre elements each of length one diameter, and the lower curves to one with lengths of 0.5 diameter. The potentials of the first and fourth elements are each equal to 1.

possible pairs of values giving the same magnifications and to deduce why they are not good combinations.

The greater range of magnification obtained with the longer central elements raises the question of further increase. However, just extending the lengths of these two elements will not help a great deal. A much better approach is to place a fifth electrode in the middle of the system, and for the moment one can think of this merely as a 'spacer' keeping the two magnification-controlling electrodes a good distance apart! Before discussing this way of obtaining zoom behaviour on a par with that obtainable with quite modest cameras, it will be helpful to examine a somewhat specialized five element cylinder lens.

4.6.2 The afocal lens

We consider in this section a *combination* of two lenses which has an interesting and useful property. Figure 4.8 shows the optics of this combination schematically and illustrates the construction, using principal rays, of the image of an object to the left of the first lens of the pair. The essential point of the lens combination is that the second focal point of the first (left hand) lens coincides with the first focal point of the second (right hand) lens. This means that the first principal ray, drawn from the top of the object parallel to the axis, passes through this common focal point and emerges from the second lens again parallel to the axis. The immediate

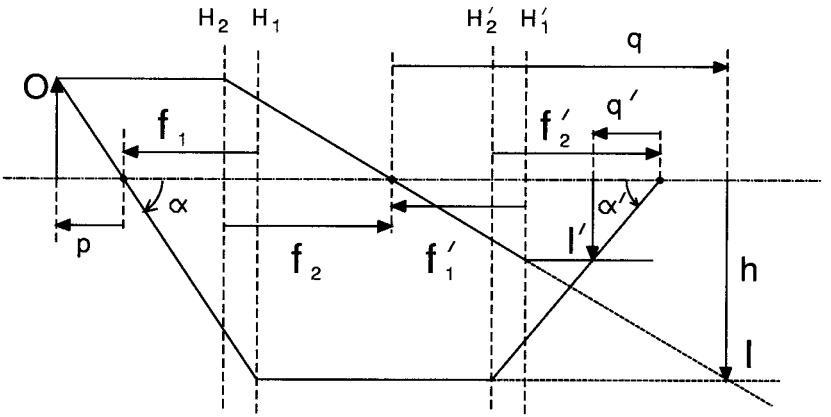

Figure 4.8 The construction of the image of a given object by two lenses having a common focal point in the region between them.

consequence of this is that, wherever the object may be, the image will have the same size: in other words, the magnification of the lens combination does not depend on the position of the object. The second principal ray is drawn through the first focal point of the left hand lens, emerges from that lens parallel to the axis and is then deviated to emerge from the right hand lens as though from the second focal point of that lens.

The focal lengths and the object and image distances are indicated on the figure and we use Newton's relation to examine their relationship. The first lens would produce an intermediate image I of O a distance q from the common focal point such that $q = f_1 f_2/p$ where p is the distance of the object from the first focal point of the left hand lens. This image lies beyond the first focal point of the second lens and is imaged by that lens as I', a distance q' from the second focal point of this lens such that $q' = f_1' f_2'/q$. In terms of the position of the original object $q' = (f_1' f_2'/f_1 f_2)p$. If the two lenses are identical then we have the interesting result that $q' = p$ and the final image is the same distance from the second focal point of the right hand lens as the object is from the first focal point of the left hand lens. The fact that rays entering parallel to the axis emerge parallel to the axis means that the lens system has no focal points—it is *afocal* or telescopic.

Remembering that F_1 is a negative quantity, the distance between the object and the final image is $2(F_2 - F_1)$, and this must be held constant, independent of the overall voltage ratio, if the special property of this lens is to be maintained. The values of F_1 and of F_2 will change, of course, and if the difference is not to change, each lens must consist of three electrodes. The last element of the left hand lens and the first element of the right hand lens are at the same potential (otherwise there would be further lens action in the middle of the system) so the entire lens consists of five elements. Because the two lenses are identical

$$\frac{V_2}{V_1} = \frac{V_4}{V_3} \qquad \text{and} \qquad \frac{V_3}{V_1} = \frac{V_5}{V_3}$$

and therefore

$$\frac{V_3}{V_1} = \left(\frac{V_5}{V_1}\right)^{1/2}.$$

The magnification of the lens can be calculated from the figure and is $M = f_1'/f_2 = f_1/f_2$ for identical lenses. The angular magnification is given by

$$M_\alpha = \frac{\alpha'}{\alpha} = \left(\frac{h}{f_2'}\right)\Big/\left(\frac{h}{f_1}\right) = \frac{f_1}{f_2'} = \frac{f_1}{f_2} = M.$$

The magnification, angular magnification and the overall voltage ratio are related by the law of Helmholtz and Lagrange

$$M M_\alpha \left(\frac{V_5}{V_1}\right)^{1/2} = 1 \qquad \text{so} \qquad M = M_\alpha = \left(\frac{V_5}{V_1}\right)^{-1/4}.$$

Apart from the trivial case of a lens which is electrically and mechanically symmetric, this is the only lens for which the magnification can be calculated without any knowledge of the potential distribution; it depends only on the *identity* of the two component lenses.

An important application of this lens is in the production of a well defined particle beam. The images of an object side window and pupil will have the same axial separation and be magnified by the same factor, so if the object side apertures are of equal diameter, the particle beam will be bounded by a cylindrical surface in a region of the image space though no apertures are present there, reducing the chance of unwanted collisions of the beam with the metalwork. The beam angle and the pencil angle will be equal in this situation.

4.6.3 The zoom afocal lens

If we remove the constraint on the two component lenses of figure 4.8, but maintain the same distance between conjugates, the lens system will not, in general, remain afocal. It does then, however, have the desirable property of the zoom lens. The magnification is controlled by the relative strengths of the first and second lenses in just the same way as is done by the two centre electrodes of a four cylinder lens. For the moment we keep the potential of the centre electrode at the geometric mean of those of the first and fifth electrodes and show in figure 4.9 (by broken curves) the values of the potentials of the second and fourth electrodes required to give a range of magnifications. The particular lens geometry consists of two lenses, each of three cylinders having lengths of 1.25, 0.5 and 1.25 diameters joined end to end to give a five element lens with a centre element 2.5 diameters long and a total length of six diameters, the same as the two four cylinder lenses shown in figure 4.7. The very substantial increase in the range of magnifications is apparent.

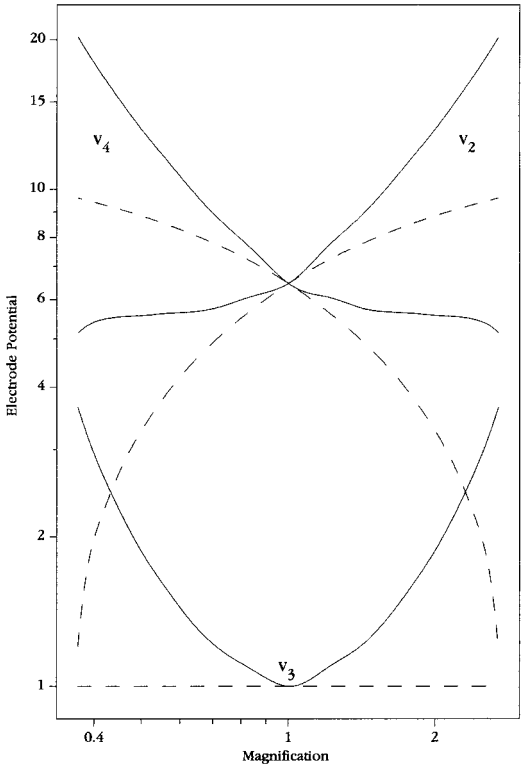

Figure 4.9 The electrode potentials required to give a range of magnifications for a five cylinder lens of total length six diameters with $V_5 = V_1 = 1$. Two sets of data are shown: the broken curves are for the case of $V_3 = 1$, and the solid curves are for V_3 adjusted to give a pupil magnification equal to the window magnification.

Only four elements are necessary for a lens to have a zoom capability over a range of voltage ratios, and the present lens has, therefore, an extra element which might be used to control another parameter of the lens. There is more than one possibility, but in the present context the control of a second conjugate pair would allow the lens to be again afocal. The effect of adjusting the potential of the centre electrode (with compensating adjustments of those of the second and fourth) is illustrated in figure 4.10 for this same six diameter length lens. The potentials have been adjusted to give a window magnification of -1 and the figure shows the magnification of a pupil located $D/2$ to the right of the window as a function of the voltage ratio V_3/V_1 for a number of values of the overall voltage ratio, V_5/V_1. The points marked • correspond to $V_3/V_1 = \sqrt{V_5/V_1}$. Notice that, for both accelerating and retarding lenses, V_3 has to be increased to make the pupil magnification equal that of the window and so make the lens afocal. For an overall

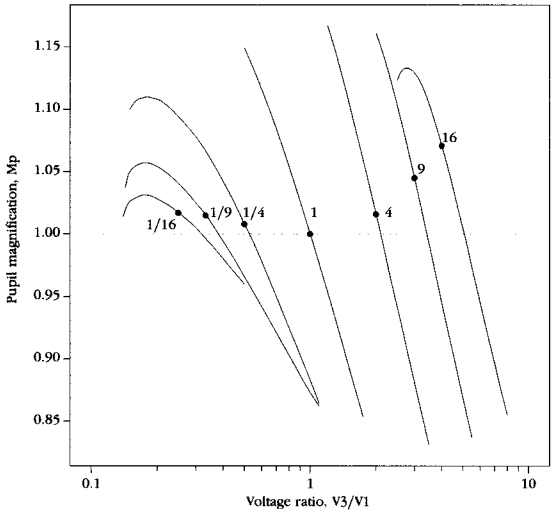

Figure 4.10 The pupil magnification of a five cylinder lens as a function of the potential of the centre element for various values of the overall voltage ratio, V_5/V_1. The lens is operated to give a window magnification of -1 and the object pupil is at a distance $D/2$ from the object window.

voltage ratio of unity, the potentials of the three centre electrodes are shown also in figure 4.9 (by full curves) for this zoom afocal configuration.

5 Aberrations

So far we have considered only the focusing of paraxial rays of charged particles homogeneous in energy. The true situation is more complex and we should consider rays which are non-paraxial and particles which have a small spread in energy. Particle lenses are subject to the same optical aberrations as photon lenses and the effects can be much worse because it is not possible to use materials of different dispersion to reduce chromatic aberration or to grind surfaces of special forms to reduce spherical and other aberrations.

5.1 Spherical aberration

The handling of particle beams is usually an axis-centred problem so spherical aberration is normally much more serious than the off-axis aberrations, such as coma, astigmatism and distortion. We shall therefore consider only rays which cross the axis at some point. Figure 5.1 is a schematic diagram showing the emergent asymptote corresponding to a ray incident from an axial object point. For meridional rays, the relationship of the radial positions and slopes of the rays at the first and second focal planes of the lens can be expressed, following Verster [6] as

$$\begin{aligned} r_2' = &-\frac{r_1}{f_2} + m_{13}r_1'^3 + m_{14}r_1'^2\frac{r_1}{f_2} + m_{15}r_1'\left(\frac{r_1}{f_2}\right)^2 \\ &+ m_{16}\left(\frac{r_1}{f_2}\right)^3 + q_{11}r_1'^5 + q_{12}r_1'^4\frac{r_1}{f_2} + q_{13}r_1'^3\left(\frac{r_1}{f_2}\right)^2 \\ &+ q_{14}r_1'^2\left(\frac{r_1}{f_2}\right)^3 + q_{15}r_1'\left(\frac{r_1}{f_2}\right)^4 + q_{16}\left(\frac{r_1}{f_2}\right)^5 + \cdots \qquad (5.1a) \end{aligned}$$

$$\begin{aligned} -\frac{r_2}{f_1} = &\, r_1' + m_{23}r_1'^3 + m_{24}r_1'^2\frac{r_1}{f_2} + m_{25}r_1'\left(\frac{r_1}{f_2}\right)^2 \\ &+ m_{26}\left(\frac{r_1}{f_2}\right)^3 + q_{21}r_1'^5 + q_{22}r_1'^4\frac{r_1}{f_2} + q_{23}r_1'^3\left(\frac{r_1}{f_2}\right)^2 \\ &+ q_{24}r_1'^2\left(\frac{r_1}{f_2}\right)^3 + q_{25}r_1'\left(\frac{r_1}{f_2}\right)^4 + q_{26}\left(\frac{r_1}{f_2}\right)^5 + \cdots \qquad (5.1b) \end{aligned}$$

TSMC EX. 1016 Page 91 of 139

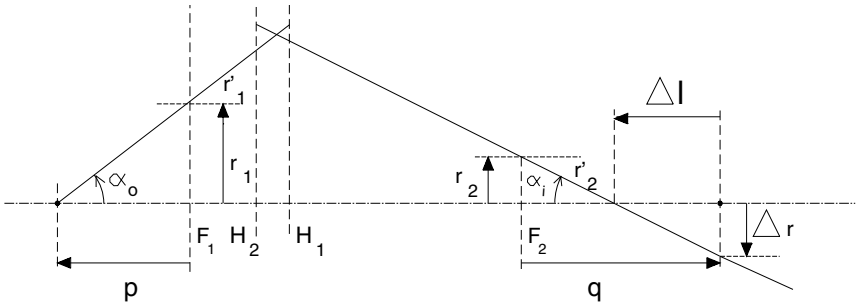

Figure 5.1 Definition of the ray parameters for the description of the meridional aberration coefficients.

remembering that f_1 is negative†. The coefficients, m_{ij} and q_{ij}, are dimensionless and also negative. The terms of fifth order are not often used, but are included here as we shall later show how some of them may be deduced from fairly simple ray traces.

The coefficient of spherical aberration for a lens operated with magnification M, $C_{s(M)}$, is defined by

$$\Delta r = -MC_{s(M)}\alpha_o^3$$

where $\alpha_o = -r_1'$. From figure 5.1 we can write $r_1 = -pr_1'$, and $\Delta r = r_2 + qr_2'$ for the distance off-axis at which the ray crosses the Gaussian image plane. The object and image distances can be written in terms of the focal lengths and the magnification as $p = -f_1/M$ and $q = -Mf_2$. Substitution of these expressions into equation (5.1), ignoring the fifth order terms, gives the following relationship between the coefficient of spherical aberration and the meridional coefficients:

$$\begin{aligned} C_{s(M)} = -m_{13}f_2 - (m_{14} + m_{23})\frac{f_1}{M} - (m_{15} + m_{24})\left(\frac{f_1}{M}\right)^2/f_2 \\ - (m_{16} + m_{25})\left(\frac{f_1}{M}\right)^3/f_2^2 - m_{26}\left(\frac{f_1}{M}\right)^4/f_2^3. \end{aligned} \tag{5.2}$$

Each term of equation (5.2) has a positive value so no cancellation is possible. C_s has the dimensions of Δr, but it is a common convention to express both in terms of the diameter of the lens, and all the values presented in this book and in the **LENSYS** program are strictly of C_s/D. Equation (5.2) shows that $C_{s(M)}$ could be expressed as a fourth order polynomial in the object distance, $p = -f_1/M$, but it is more usual to write it in terms of the magnification, M,

$$C_{s(M)} = C_{s_0} + C_{s_1}/M + C_{s_2}/M^2 + C_{s_3}/M^3 + C_{s_4}/M^4 \tag{5.3}$$

where the C_{s_i} are properties of the lens itself.

† A number of the equations in this chapter differ from their counterparts in other texts as a consequence of our Cartesian sign convention.

There are certain well established relationships between some of the eight meridional aberration coefficients

$$m_{14} = 3m_{23} + 1.5 \qquad m_{15} = m_{24} \qquad m_{25} = 3m_{16} + 1.5 \tag{5.4}$$

and so the five spherical aberration coefficients are

$$\begin{aligned}
C_{s_0} &= -m_{13} f_2 \\
C_{s_1} &= -(m_{14} + m_{23}) f_1 = -(4m_{23} + 1.5) f_1 \\
C_{s_2} &= -(m_{15} + m_{24}) f_1^2/f_2 = -2m_{24} f_1^2/f_2 \\
C_{s_3} &= -(m_{16} + m_{25}) f_1^3/f_2^2 = -(4m_{16} + 1.5) f_1^3/f_2^2 \\
C_{s_4} &= -m_{26} f_1^4/f_2^3.
\end{aligned} \tag{5.5}$$

If the object is at infinity, $r_1' = 0$ and the magnification is therefore zero. It is then necessary to use equation (5.1*b*) directly. Including the relevant fifth order term, $q_{26}(r_1/f_2)^5$, we have

$$\Delta r = r_2 = -m_{26} f_1 \left(\frac{r_1}{f_2}\right)^3 - q_{26} f_1 \left(\frac{r_1}{f_2}\right)^5.$$

The third order term can be written as

$$\Delta r = C_{s_4} \left(\frac{r_1}{f_1}\right)^3. \tag{5.6}$$

5.1.1 *The axial displacement*

The aberrated ray crosses the axis a distance Δl before the Gaussian image point where

$$\begin{aligned}
\Delta l = -\frac{\Delta r}{r_2'} = -MC_s \frac{\alpha_o^3}{\alpha_i} &= -\frac{M}{M_\alpha} C_s \alpha_o^2 \\
&= M^2 C_s \alpha_o^2 \frac{f_2}{f_1} = M^4 C_s \alpha_i^2 \left(\frac{f_2}{f_1}\right)^3
\end{aligned} \tag{5.7}$$

and where $\alpha_i = -r_2'$ is the beam angle in the image space.
For parallel input $M = 0$ so $M^4 C_s = C_{s_4}$ and, using $\alpha_i = r_1/f_2$,

$$\Delta l = \left(\frac{f_2}{f_1}\right)^3 C_{s_4} \alpha_i^2 \quad = \frac{f_2}{f_1} C_{s_4} \left(\frac{r_1}{f_1}\right)^2. \tag{5.8}$$

If we wish to include the fifth order term without excessive algebra, we can write

$$r_2' = -r_2/\Delta l = -r_1/(f_2 + \Delta l)$$

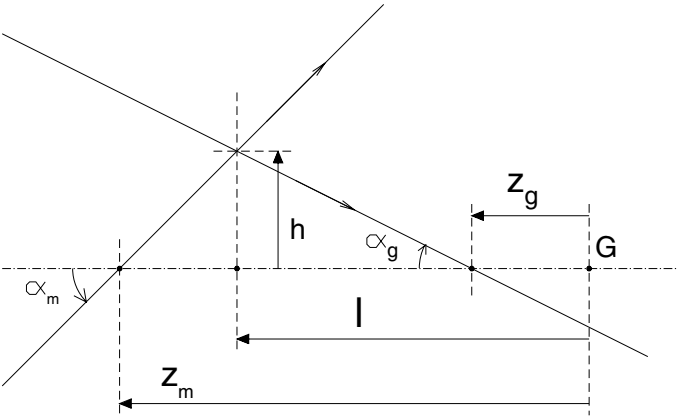

Figure 5.2 Aberrated rays in the region of the disc of minimum confusion.

giving

$$r_2 = \frac{r_1 \Delta l}{f_2 + \Delta l}$$

from which we find the following working expression:

$$G(\Delta l) = \left(\frac{f_2}{f_1}\right)\frac{\Delta l}{f_2 + \Delta l} = -m_{26}\left(\frac{r_1}{f_2}\right)^2 - q_{26}\left(\frac{r_1}{f_2}\right)^4 \tag{5.9}$$

which we shall use later.

5.1.2 The disc of minimum confusion

If the object is a point source of particles, the 'image' in the Gaussian plane will be a round spot with a radius determined by the pencil angle and the aberration coefficients of the lens. Rays which have passed close to the limits of the entrance pupil will contribute to the outer parts of this spot, having crossed the axis somewhat closer to the lens. Between the Gaussian plane and the point on the axis where these marginal rays are focused, there will be some rays which cross the axis going upwards and some going downwards, with the former being more steeply inclined, and the spot formed by the envelope of all the rays will have a minimum radius somewhere between these points. This spot is known as 'the disc of minimum confusion'.

Figure 5.2 shows the marginal ray and a general ray which intersect a distance h from the axis; the axial position of the intersection is a distance l from the Gaussian focus, G. We can express h in two ways

$$h = \alpha_g(z_g - l) = \alpha_m(z_m - l)$$
$$A\alpha_g^3 - l\alpha_g = A\alpha_m^3 - l\alpha_m$$

so

$$l = A(\alpha_m^2 + \alpha_m\alpha_g + \alpha_g^2)$$

and

$$h = -A(\alpha_m^2\alpha_g + \alpha_m\alpha_g^2)$$

where $A = C_{s4}\left(f_2/f_1\right)^3$ by equation (5.8), and is negative. We find the minimum value of h by differentiation

$$\frac{\mathrm{d}h}{\mathrm{d}\alpha_g} = -A(\alpha_m^2 + 2\alpha_m\alpha_g) = 0 \quad \text{if} \quad \alpha_g = -\frac{\alpha_m}{2}$$

and this leads to the radius of the disc of minimum confusion as

$$h_{min} = A\frac{\alpha_m^3}{4}$$

and its position as

$$l = \frac{3}{4}A\alpha_m^2.$$

The radius is a quarter of that of the Gaussian plane spot and it is three quarters of the way from the Gaussian image to the most aberrated image point.

5.2 Chromatic aberration

Figure 5.3 illustrates the passage of two pairs of rays through a lens system ignoring the effects of spherical aberration. One ray of each pair has the correct energy to form an image at I. The other has a slightly higher energy. We consider first the pair of rays which start from a point on the axis. The 'correct' ray crosses the axis at the image, but the higher energy ray crosses the axis a distance δl beyond this and, at the image plane, is a distance δr above the axis. These distances are related by the angle, α_i in the image space and can be referred to the corresponding angle at the object using the Helmholtz–Lagrange relation.

$$\delta r = \alpha_i\delta l = M_\alpha\alpha_o\delta l = \frac{1}{M}\left(\frac{V_o}{V_i}\right)^{\frac{1}{2}}\alpha_o\delta l.$$

A *coefficient of chromatic aberration*, C_c, has historically been defined by $\delta r = -MC_c\alpha_o\frac{\delta V}{V_o}$ and we may, therefore, express this coefficient in terms of either observable by

$$C_c = \frac{1}{M\alpha_o}\frac{\delta r}{(\delta V/V_o)} \tag{5.10a}$$

$$= \frac{1}{M^2}\left(\frac{V_o}{V_i}\right)^{\frac{1}{2}}\frac{\delta l}{(\delta V/V_o)}. \tag{5.10b}$$

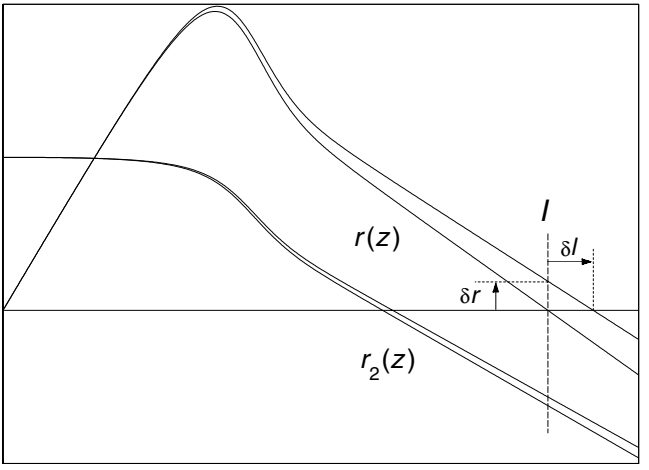

Figure 5.3 Two pairs of rays traced to illustrate chromatic aberration. The lower ray of each pair has the correct energy to form an image at I and the other rays, which have a slightly higher energy, cross the image plane at points which are displaced from the true image points.

The second of these equations illustrates the non-dependence of the aberration coefficient on the angle α_o.

This coefficient is properly referred to as the coefficient of *axial* chromatic aberration to distinguish it from the effect illustrated by the other pair of rays in figure 5.3 which are launched parallel to the axis from a point above the axis. The 'correct' ray crosses the image plane and defines the magnification of the lens. The higher energy ray crosses the image plane closer to the axis, with a lower magnification. This effect is described as *chromatic aberration of magnification* and the coefficient, C_D, which quantifies the effect is sometimes referred to as the coefficient of *lateral* chromatic aberration. It is defined in terms of the radial change, δr_I, in the ray at the image plane by the following equations.

$$\delta r_I = -MC_D r_o \left(\frac{\delta V}{V_o}\right)$$
$$C_D = \frac{-1}{Mr_o}\frac{\delta r}{(\delta V/V_o)} \tag{5.11}$$

where r_o is the initial off-axis distance of the ray and, for the case of the general ray, $r_2(z)$, which crosses the axis at an angle of 45°, is equal to f_1.

This aberration is an off-axis effect and we discuss, in the next section, the off-axis effects due to the angle dependent behaviour.

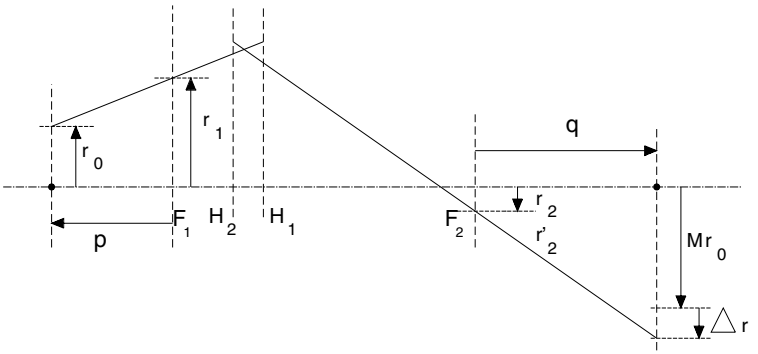

Figure 5.4 Definition of the ray parameters for the description of the off-axis aberrations of a lens.

5.3 Off-axis aberrations

Figure 5.4 is similar to figure 5.1, but indicates the asymptotes to a ray from an object point, a distance r_0 from the axis, which crosses the first focal plane at a radial distance r_1. The path of the ray is still in a plane containing the axis; we shall not be considering skew rays. Using equations (5.1) again, we write the position of the ray at the image plane as the sum of the geometric, Gaussian, distance, Mr_0, and an aberration term which can be written as the sum of terms involving r_0^3, $r_0^2 r_1$, $r_0 r_1^2$ and r_1^3.

$$
\begin{aligned}
\Delta r = Mr_0 &+ r_0^3\left[M^4 m_{13}\frac{f_2}{f_1^3} + M^3 m_{23}\frac{1}{f_1^2}\right] \\
&- r_0^2 r_1\left[3M^4 m_{13}\frac{f_2}{f_1^3} + M^3(m_{14}+3m_{23})\frac{1}{f_1^2} + M^2 m_{24}\frac{1}{f_1 f_2}\right] \\
&+ r_0 r_1^2\left[3M^4 m_{13}\frac{f_2}{f_1^3} + M^3(2m_{14}+3m_{23})\frac{1}{f_1^2}\right. \\
&\qquad \left. + M^2(m_{15}+2m_{24})\frac{1}{f_1 f_2} + Mm_{25}\frac{1}{f_2^2}\right] \\
&- r_1^3\left[M^4 m_{13}\frac{f_2}{f_1^3} + M^3(m_{14}+m_{23})\frac{1}{f_1^2} + M^2(m_{15}+m_{24})\frac{1}{f_1 f_2}\right. \\
&\qquad \left. + M(m_{16}+m_{25})\frac{1}{f_2^2} + m_{26}\frac{f_1}{f_2^3}\right].
\end{aligned}
$$

The terms in r_0^3 describe the distortion of the image, because they imply a magnification which depends on the position of the point in the object but not on the lens aperture. From equation (5.5) we see that C_{s_0} and C_{s_1} describe this aberration. The terms in $r_0^2 r_1$ describe the combined effects of astigmatism and curvature of field and those in $r_0 r_1^2$ describe coma. The terms in r_1^3 describe spherical aberration

and transform to the form of equation (5.2) on writing $r_1 = -\frac{f_1}{M}\alpha_o$. If $r_0 = r_1$, which implies $M = 0$, only the term in m_{26}, equivalent to C_{s_4}, remains and equation (5.6) is recovered. Notice that m_{26} appears only in the coefficient of r_1^3.

While the terms in $r_0^2 r_1$ determine the curvature of field for a specific magnification—and the curvature is usually dominated by the effect of astigmatism—there is a limiting value which obtains even in the absence of astigmatism. This minimum curvature is expressed by the *Petzval integral*, which is never zero for a particle lens. The integral can only be zero if there are contributions from divergent lens elements and, as we saw in chapter 2, immersion lenses, with which we are exclusively concerned, are always convergent. It is a property of the given lens and depends only on the potential distribution and not on the choice of conjugates. It is, therefore, most readily calculated along with the basic lens parameters.

5.4 Interrelations of the spherical aberration coefficients

By considering a lens to be 'thin' in the sense that R does not change over the distance for which T is significant, it is possible to factor a term out of the aberration integral and hence develop simple relationships between the m_{ij} coefficients [17]. It has been demonstrated [16] that these remain reasonably valid even when the 'thin' approximation does not appear to be appropriate. The relationships are of the form $m_{2j}/m_{1j} = \sigma$ where the value of the ratio has been determined empirically as $\sigma = (V_2/V_1)^{1/4}$ for a two cylinder lens. Taken in conjunction with the well established relationships between three pairs of m_{ij} (equation (5.4)) and ignoring the added '1.5' terms, which is a reasonable approximation except for very strong lenses, there appears to be a pattern of interrelations between all the m_{ij} coefficients,

$$m_{i6} \simeq \sigma m_{i5}/3 \simeq \sigma^2 m_{i4}/3 \simeq \sigma^3 m_{i3} \qquad m_{2j} \simeq \sigma m_{1j}.$$

With these factors, equation (5.2) becomes

$$C_{s(M)} = -m_{13} f_2 \Big[1 + 4\frac{\sigma}{M}\frac{f_1}{f_2} + 6\Big(\frac{\sigma}{M}\Big)^2\Big(\frac{f_1}{f_2}\Big)^2 + 4\Big(\frac{\sigma}{M}\Big)^3\Big(\frac{f_1}{f_2}\Big)^3 + \Big(\frac{\sigma}{M}\Big)^4\Big(\frac{f_1}{f_2}\Big)^4\Big]. \tag{5.12}$$

If σ is indeed $(V_2/V_1)^{1/4}$, then $\sigma^2 = -f_2/f_1$ so $\sigma f_1/f_2 = -1/\sigma$ and, writing C_{s_0} for $-m_{13} f_2$, we can express the spherical aberration coefficient for a given magnification as

$$C_{s(M)} = C_{s_0}\left(1 - \frac{1}{\sigma M}\right)^4. \tag{5.13}$$

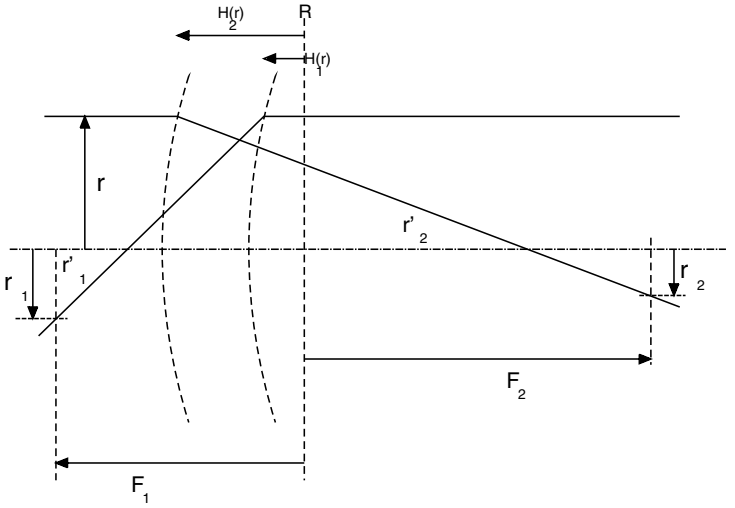

Figure 5.5 Diagram to illustrate the curved principal surfaces.

A careful study of more complex lenses shows that this is indeed rather approximate, but another result based on the same relationships does seem to have a wider validity. We define a function, Y, by

$$Y = C_{s_0}\left(\frac{-\sigma f_1}{f_2}\right)^4 + C_{s_1}\left(\frac{-\sigma f_1}{f_2}\right)^3 + C_{s_2}\left(\frac{-\sigma f_1}{f_2}\right)^2 + C_{s_3}\left(\frac{-\sigma f_1}{f_2}\right) + C_{s_4} \tag{5.14}$$

and substitute for the C_{s_i} the appropriate terms of equation (5.12), giving

$$Y = C_{s_0}\left(\frac{-\sigma f_1}{f_2}\right)^4 (1 - 4 + 6 - 4 + 1) = 0.$$

If we write Y_+ to represent the sum of the first, third and fifth terms of equation (5.14) and Y_- to represent the sum of the second and fourth (which is usually negative), then the ratio $\frac{Y_+ + Y_-}{Y_+ - Y_-}$ will be a fair measure of the accuracy of the statement $Y = 0$. For a wide range of lenses (and the reader is invited to test this using **LENSYS**) this ratio is less than 0.01 and usually much less.

5.4.1 The principal surfaces

We noted in chapter 1 that the principal 'planes' were not in fact flat and we are now in a position to examine their true form. Figure 5.5 shows the principal surfaces as defined by the intersection of the object and image asymptotes of rays incident at a radial distance, r, parallel to the axis. We can express the distance from the second *Gaussian* focal point to the projection of the intersection on to the axis, $H_2(r)$, as

$$F_2 - H_2(r) = -(r_2 + r)/r_2'$$

using equations (5.1)

$$F_2 - H_2(r) = f_2 + (f_1 m_{26} + f_2 m_{16})\left(\frac{r}{f_2}\right)^2$$

and writing $\sigma^2 = -f_2/f_1$

$$F_2 - H_2(r) = f_2\left[1 + m_{13}\sigma^2(\sigma - 1)\left(\frac{r}{f_2}\right)^2\right].$$

A similar analysis gives, for the first principal surface,

$$\begin{aligned} F_1 - H_1(r) &= -(r_1 + r)/r_1' \\ &= f_1 + (f_1 m_{23} + f_2 m_{13})\left(\frac{r}{f_1}\right)^2 \\ &= f_1\left[1 - m_{13}\sigma(\sigma - 1)\left(\frac{r}{f_1}\right)^2\right]. \end{aligned}$$

Note that m_{13} is negative and $\sigma > 1$ for an accelerating lens, so both principal surfaces are concave towards the high potential side of the lens.

5.5 The determination of aberration coefficients

If the potential distribution throughout the lens is known with sufficient accuracy, rays will follow their correct paths and, if they are not paraxial, will show the effects of the lens aberrations. In principle, then, all that is needed is careful ray tracing to find the intercept of the ray at the Gaussian image plane. This raises the first problem: where is the Gaussian image plane? It is necessary to trace more than one ray and extrapolate to zero α_o or r_1 as appropriate and it is tempting to use the axial displacement directly. If only the axial potential is known, then calculations must be based on the effect of the terms omitted from the axial equation of motion.

5.5.1 Direct ray tracing

Figure 5.6 shows rays traced in an accelerating and a decelerating lens.
The much greater effect of spherical aberration in the retarding lens is apparent. To illustrate a procedure for the determination of the aberration coefficients appropriate to parallel input rays we shall analyse some data obtained with the 'student' version of the CPO-2D program.

For the purposes of the present analysis, the important parameters of the trajectory are those indicated in figure 5.7. The position of the focal point, F_2, is not known *a priori* and we cannot, therefore, determine r_2 directly, but,from measurements made at the reference plane, the program provides values for the axial position, z_i, at which a ray launched parallel to the axis at a radial distance r_1 crosses the axis and the slope, r_2', of the ray at that point. From these values

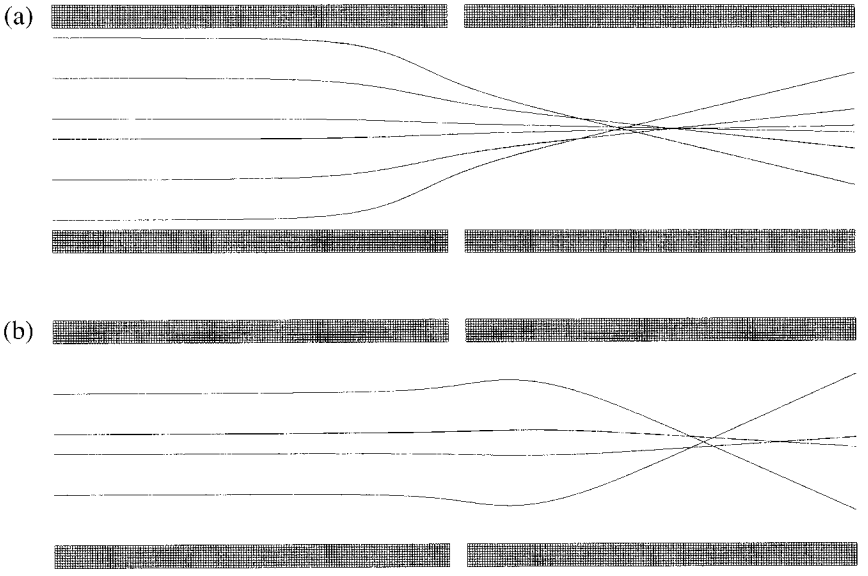

Figure 5.6 Rays traced in a two cylinder lens having voltage ratios of (a) 1 : 10, and (b) 10 : 1. The aberrations of the retarding lens are so great that the outermost ray of (a) would strike the inner surface of the electrode well before the end of the diagram.

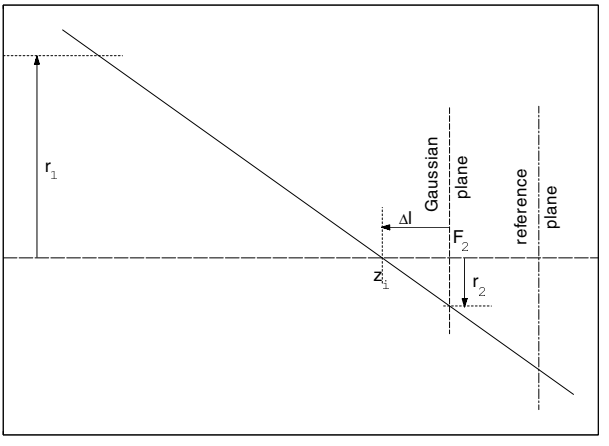

Figure 5.7 The path of an aberrated ray which crosses the Gaussian plane at a distance r_2 having crossed the axis at a position z_i a distance Δl before.

we can determine the focal parameters f_2 and F_2, and the aberration coefficients m_{16}, m_{26}, q_{16} and q_{26}. If we trace rays in a second lens, having the reciprocal

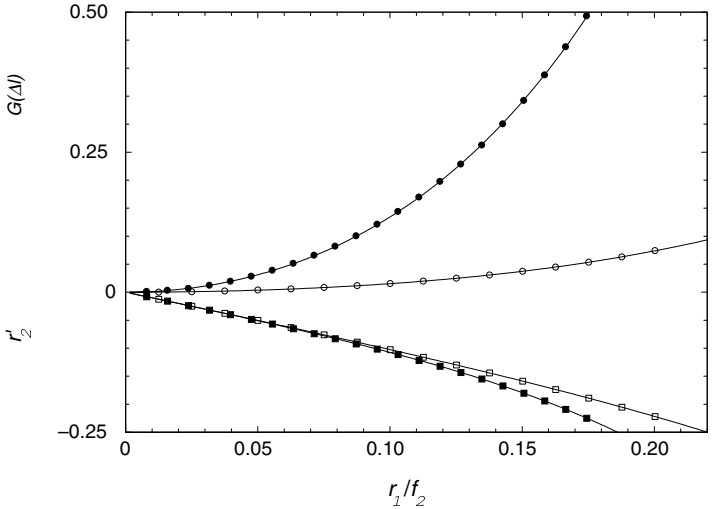

Figure 5.8 Data plots used to determine the spherical aberration coefficients of the lenses shown in figure 5.6. The curves represent $r_2' = -\frac{r_1}{f_2} + m_{16}(\frac{r_1}{f_2})^3 + q_{16}(\frac{r_1}{f_2})^5$ and $G(\Delta l) = m_{26}(\frac{r_1}{f_2})^2 + q_{26}(\frac{r_1}{f_2})^4$. The solid data points refer to the accelerating lens and the open data points to the decelerating lens.

voltage ratio, we also find f_1, F_1, m_{13}, m_{23}, q_{11} and q_{21}.

The first step is to find F_2 by writing $z_i = F_2 + \Delta l$, where Δl is negative and is given by equation (5.8). A graph of z_i against r_1^2 is almost a straight line and a quadratic fit will give a very good value for F_2. We next plot r_2' against r_1 and fit an expression in odd powers of r_1. From equation (5.1*a*) the coefficient of the first term is $-1/f_2$ and those of the other terms will give the aberration coefficients. It is a little neater to repeat this with r_1/f_2 as the abscissa and we show such a plot in figure 5.8. Now that we have a value for f_2 it is possible to use equation (5.9), remembering that the ratio $f_2/f_1 = -(V_2/V_1)^{1/2}$, to find the remaining coefficients. In figure 5.8 we show the results for both the accelerating (1 : 10) and decelerating (10 : 1) lenses. The values of the coefficients are found to be

1 : 10	10 : 1
$m_{16} = -6.41$	-2.27
$m_{26} = -12.1$	-1.42
$q_{16} = -100$	-10.8
$q_{26} = -134$	-10.6

Figure 5.9 shows values of the chromatic aberration coefficient, C_c, determined by measuring the image plane 'miss distance', δr_I, for a ray launched with an

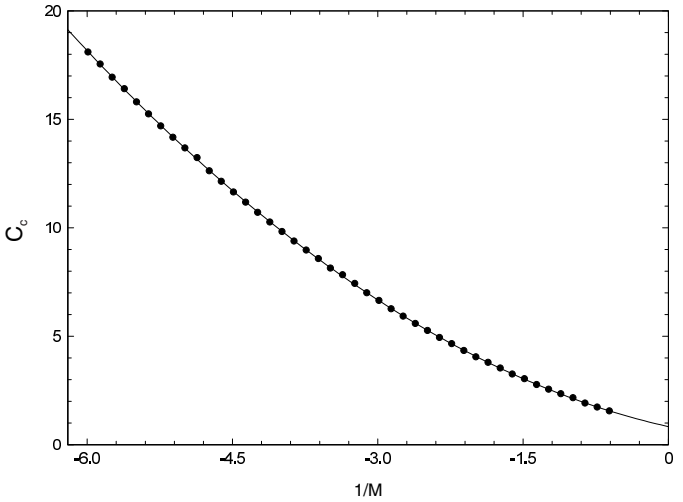

Figure 5.9 The coefficient of chromatic aberration for a two cylinder lens with a voltage ratio of 1 : 10. The data can be represented by $C_c(M) = 0.834 - 1.005/M + 0.314/M^2$.

energy 2% greater than nominal in a two cylinder lens with a potential ratio of 1 : 10. The coefficients of a power law fit to the data agree very well with values calculated directly using expressions which we develop later.

5.5.2 Perturbation of the paraxial solution

We can modify the equation of motion to take account of terms omitted in our earlier analysis or to allow for a spread in energy and, provided that the effects are small, we can write the solution to the modified equations as the simple sum of the paraxial solution and an additional term. The details of the calculation depend on the problem and are expanded in the next two sections.

5.5.2.1 Spherical aberration. In chapter 2, the higher order terms in the expansion of the axial potential were ignored as was the contribution made to the total energy of the particle by its radial velocity. These simplifications led to the paraxial equation of motion, equation (2.3), and then to the Picht equation, (2.4). In order to describe the non-paraxial behaviour, it is necessary to replace these terms. The equation of radial motion becomes

$$m\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = eE_r = e\left[\frac{r}{2}V'' - \frac{r^3}{16}V^{(4)}\right]$$

and the total energy of the particle has to be written as

$$\frac{1}{2}m\left[\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2\right] + e\left[V - \frac{r^2}{2}V''\right] = 0.$$

The equation of motion is then

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{V}\frac{\mathrm{d}r}{\mathrm{d}z} + \frac{r}{4}\frac{V''}{V} = \Big[\frac{1}{32}\frac{V^{(4)}}{V} - \frac{1}{16}\Big(\frac{V''}{V}\Big)^2\Big]r^3 + \frac{1}{8}\Big(\frac{V'''}{V} - \frac{V''V'}{V^2}\Big)r^2 r' - \frac{1}{4}\frac{V''}{V}rr'^2 - \frac{1}{2}\frac{V'}{V}r'^3. \quad (5.15)$$

To solve this equation, we write $r = r_p + r_a$ where r_p is the solution to the equation of paraxial motion and r_a is a *small* perturbation representing the aberration due to the use of better approximations to the real potential. r_a is small enough that only first order terms need to be retained. Each term of the left hand side of equation (5.15) splits into one containing r_p and one containing r_a. Those in r_p sum to zero, because they are just the terms of the paraxial equation. The terms on the right hand side involve the third power of r and so only the r_p part survives, giving an equation for the perturbation, r_a,

$$\frac{\mathrm{d}^2 r_a}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{V}\frac{\mathrm{d}r_a}{\mathrm{d}z} + \frac{r_a}{4}\frac{V''}{V} = \Big[\frac{1}{32}\frac{V^{(4)}}{V} - \frac{1}{16}\Big(\frac{V''}{V}\Big)^2\Big]r_p^3 + \frac{1}{8}\Big(\frac{V'''}{V} - \frac{V''V'}{V^2}\Big)r_p^2 r_p' - \frac{1}{4}\frac{V''}{V}r_p r_p'^2 - \frac{1}{2}\frac{V'}{V}r_p'^3. \quad (5.16)$$

Solutions of the homogeneous equation, obtained by setting the right hand side of equation (5.16) equal to zero, are known: they are just the general solutions of the Gaussian ray equation discussed in chapter 1. We can therefore write the perturbation as $\zeta(z)r_1(z) + \xi(z)r_2(z)$ where we stress the variability of ζ and ξ, reflecting the non-zero right hand side. The problem reduces to the calculation of $\xi(z_i)$ where z_i is the position of the Gaussian image. After a lot of algebra, two integrations by parts and transformation of the radial variable to the reduced radius, $R = rV^{1/4}$, the coefficient of spherical aberration can be written as

$$C_{s(M)} = \frac{1}{64V_0^{1/2}}\int_{-\infty}^{\infty}\frac{R^4}{V^{1/2}}\Big[\Big(3T^4 + 5T'^2 - \frac{11}{2}T^2T'\Big) + 4TT'\frac{R'}{R}\Big]\mathrm{d}z \quad (5.17)$$

where V_0 is the potential at the lower limit of integration†. In this equation, $R(z)$ corresponds to the ray which passes through the axial object point with unit slope. Every variable, bar one, in this equation will have been determined in the course of calculating the paraxial image position. The missing variable is T' and it is a trivial matter to find this. We noted earlier that the sign of T is not important to the calculation of the paraxial focus as it appears as T^2 in the Picht equation, but the presence of T' here shows that it does matter for the calculation of the aberrations.

† The use of the reduced radius has transformed a term in $(V/V_0)^{1/2}$ into $(1/V_0V)^{1/2}$. Almost always, the potentials will be expressed in terms of V_0.

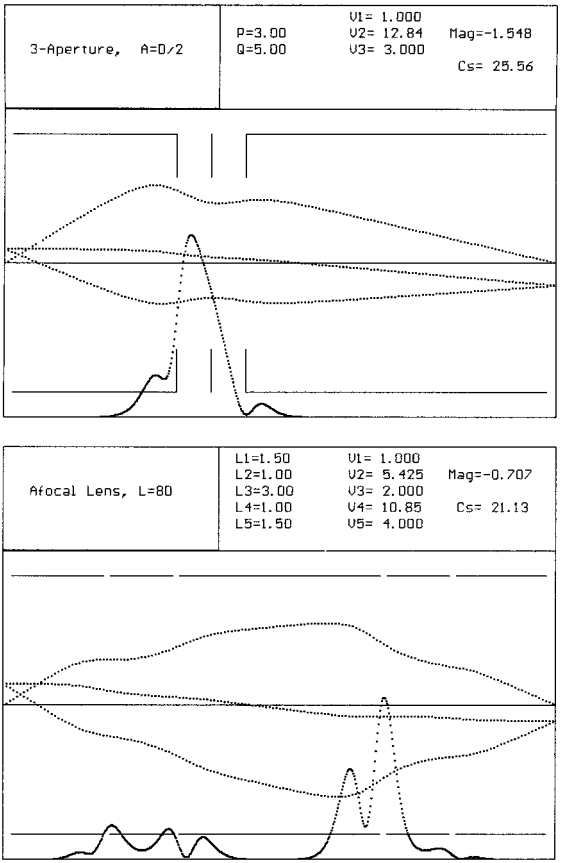

Figure 5.10 The integrands of the aberration integrals of equation (5.17) for three aperture and Afocal8 lenses.

Though the limits of integration are marked as $\pm\infty$, in practice a much shorter interval will suffice.

Figure 5.10 shows the integrand of equation (5.17) for the cases of a three aperture lens, used asymmetrically, and a five cylinder lens used in the afocal mode. The range of axial position over which the integrand is significantly greater than zero is appreciably less than the distance between the conjugate points. For a two element lens, we might expect the integrand to show two maxima, one where R is large and the other where R' is large. In the three element lens, especially with aperture electrodes, two of the four possible maxima run into each other, but the cylinder lens shows two maxima associated with each gap. The ray paths shown in this figure represent the true radius, $r(z)$.

The ratio of the reduced and true radii has the same value at any point for all trajectories, so we can write $R(z) = R_1(z) + \frac{1}{M}R_2(z)$ and expand the radial terms

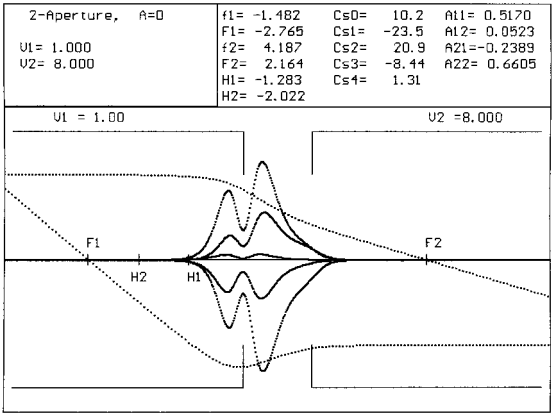

Figure 5.11 The integrands of the integrals analogous to equation (5.17) for the five spherical aberration coefficients for a two aperture lens. The curves can readily be associated with particular coefficients as these are markedly different in magnitude.

as

$$R^4 = R_1^4 + 4R_1^3R_2/M + 6R_1^2R_2^2/M^2 + 4R_1R_2^3/M^3 + R_2^4/M^4$$

and similarly for R^3R'. This allows us to separate the integral of equation (5.17) into the sum of five integrals, one for each of the C_{s_i}. Each integral is written as the product of terms involving the T-parameters,

$$c_1 = \frac{1}{V^{1/2}}\left[3T^4 + 5T'^2 - \frac{11}{2}T^2T'\right]$$
$$c_2 = \frac{4TT'}{V^{1/2}}$$

which, being functions of only the electrode geometry and potentials, are the same for all applications of the lens, and terms specific to each coefficient

$$\begin{aligned}
I_0 &= c_1R_1^4 + c_2R_1^3R_1' \\
I_1 &= c_1 4R_1^3R_2 + c_2(3R_1^2R_2R_1' + R_1^3R_2' \\
I_2 &= c_1 6R_1^2R_2^2 + c_2(3R_1R_2^2R_1' + 3R_1^2R_2R_2') \\
I_3 &= c_1 4R_1R_2^3 + c_2(R_2^3R_1' + 3R_1R_2^2R_2') \\
I_4 &= c_1R_2^4 + c_2R_2^3R_2'.
\end{aligned} \tag{5.18}$$

r_1 and r_2 are asymptotic to $\pm f_1$, which is not known *a priori* so the focal properties must be calculated using arbitrary horizontal asymptotes (usually 1) and then scaled to provide the values required for the aberration calculation. However,

to calculate the coefficient $C_s(M)$ between conjugates, the ray $R(z)$ can be used with no further scaling. Figure 5.11† shows the integrands used to calculate the five C_{s_i}. The individual curves can easily be associated with particular coefficients and for this lens they are, from the top, C_{s_2}, C_{s_0}, C_{s_4}, C_{s_3}, C_{s_1}.

5.5.2.2 The Petzval integral. The Petzval sum describes the curvature of the image surface in the absence of astigmatism. In photon optics, where one has sharp boundaries separating regions of different refractive index, the radius of curvature can be expressed in terms of the radii of the boundaries and the refractive indices and, using equations (1.3) to introduce the focal lengths of each boundary, reduces to $1/r_p = \sum(1/n_j f_j)$. The surface is a paraboloid of revolution and is concave towards the lens elements if $\sum(1/n_j f_j)$ is positive.

This will be the case for a single lens, but it is quite possible for the Petzval sum to be zero for discrete lens combinations. Indeed there is a well known photographic lens, the *Petzval lens*, which has a slight telephoto action as it consists of a convergent element followed by a divergent element. Notice that the details of the ray path make no contribution to the sum.

For charged particle optics we have the Petzval integral, a direct analogue, in which the square root of the axial potential has the rôle of the refractive index and T' accounts for the bending of the rays. The full expression is

$$Ptz = \frac{1}{8}\int_{-\infty}^{\infty}\left(\frac{V_o}{V}\right)^{1/2} T'\,\mathrm{d}z. \tag{5.19}$$

Some examples of the integrand are shown in the appendix and, though the changes of sign are apparent, the integral as a whole is never negative.

5.5.2.3 Chromatic aberration. In this case, the perturbing factor can be expressed as a constant change in the potential, and we can write the equation of motion as

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{(V+\delta V)}\frac{\mathrm{d}r}{\mathrm{d}z} + \frac{r}{4}\frac{V''}{(V+\delta V)} = 0$$

and, as δV is small in comparison with V, we can multiply top and bottom by $V - \delta V$ to give

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{V}\frac{\mathrm{d}r}{\mathrm{d}z} + \frac{r}{4}\frac{V''}{V} = \delta V\left[\frac{1}{2}\frac{V'}{V^2}\frac{\mathrm{d}r}{\mathrm{d}z} + \frac{1}{r}\frac{V''}{V^2}\right].$$

We continue, as before, by writing $r = r_p + r_c$ where r_c is a change in the trajectory, small enough that the product $r_c \delta V$ can be neglected.

$$\frac{\mathrm{d}^2 r_c}{\mathrm{d}z^2} + \frac{1}{2}\frac{V'}{V}\frac{\mathrm{d}r_c}{\mathrm{d}z} + \frac{r_c}{4}\frac{V''}{V} = \delta V\left[\frac{1}{2}\frac{V'}{V^2}\frac{\mathrm{d}r_p}{\mathrm{d}z} + \frac{r_p}{4}\frac{V''}{V^2}\right].$$

† Figures 5.10 and 5.11 have been prepared with a version of LENSYS modified to print larger dots.

We again write the perturbation as $\zeta(z)r_1(z) + \xi(z)r_2(z)$ where r_1 and r_2 are the general solutions of the ray equation. In this case we are interested in the values of both $\zeta(z_i)$ and $\xi(z_i)$

Using the expressions for the chromatic aberration coefficients (equations (5.10) and (5.11)) and then changing the radial variable to the reduced radius we find

$$\begin{aligned} C_c &= \frac{1}{V_0^{1/2}} \int_{-\infty}^{\infty} \frac{3}{8} \frac{V'^2}{V^{5/2}} r^2 \, \mathrm{d}z \\ &= \frac{1}{V_0^{1/2}} \int_{-\infty}^{\infty} \frac{3}{8} \frac{V'^2}{V^2} R^2 \, \mathrm{d}z \\ &= \frac{1}{V_0^{1/2}} \int_{-\infty}^{\infty} \frac{3}{8} T^2 R^2 \, \mathrm{d}z \end{aligned} \tag{5.20}$$

and

$$\begin{aligned} C_D &= \frac{1}{V_0^{1/2} f_1} \int_{-\infty}^{\infty} \frac{3}{8} \frac{V'^2}{V^{5/2}} r_2 r \, \mathrm{d}z \\ &= \frac{1}{V_0^{1/2} f_1} \int_{-\infty}^{\infty} \frac{3}{8} \frac{V'^2}{V^2} R_2 R \, \mathrm{d}z \\ &= \frac{1}{V_0^{1/2} f_1} \int_{-\infty}^{\infty} \frac{3}{8} T^2 R_2 R \, \mathrm{d}z. \end{aligned} \tag{5.21}$$

We have also substituted T for the ratio of the potential gradient to the potential.

In these expressions $R(z)$ is the ray which passes through the axial object point with unit slope and we can expand it in terms of the general solutions by $R(z) = R_1(z) + (1/M)R_2(z)$. This enables us to find the magnification dependence of the chromatic aberration coefficients. As in the case of spherical aberration, these coefficients have to be calculated using arbitrarily normalized representations of r_1 and r_2 and then scaled by multiplying by f_1^2. Only in the calculation of $C_c(M)$ is this scaling not necessary, because $R(z)$ is the correctly normalized reduced ray.

$$\begin{aligned} C_c &= \frac{1}{V_0^{1/2}} \int_{-\infty}^{\infty} \frac{3}{8} T^2 \left(R_1^2 + \frac{2}{M} R_1 R_2 + \frac{1}{M^2} R_2^2 \right) \mathrm{d}z \\ &= C_{c_0} + C_{c_1}/M + C_{c_2}/M^2 \end{aligned} \tag{5.22}$$

$$\begin{aligned} C_D &= \frac{1}{V_0^{1/2} f_1} \int_{-\infty}^{\infty} \frac{3}{8} T^2 \left(R_1 R_2 + \frac{1}{M} R_2^2 \right) \mathrm{d}z \\ &= C_{D_0} + C_{D_1}/M. \end{aligned} \tag{5.23}$$

It follows from these equations that $C_{D_0} = C_{c_1}/2f_1$ and $C_{D_1} = -C_{c_2}/f1$. Note that both C_{c_1} and C_{D_1} are negative quantities.

5.6 The aberrations of retarding lenses

5.6.1 Spherical aberration

The equations (5.1) which describe the ray position and slope at the second focal plane in terms of the values of these parameters at the first may be inverted algebraically to describe the effect of reversing the direction of the rays

$$\begin{aligned} r_1' &= -\frac{r_2}{f_1} + m_{26}r_2'^3 + m_{25}r_2'^2\frac{r_2}{f_1} + m_{24}r_2'\left(\frac{r_2}{f_1}\right)^2 + m_{23}\left(\frac{r_2}{f_1}\right)^3 + \cdots \\ \frac{r_1}{f_2} &= -r_2' - m_{16}r_2'^3 - m_{15}r_2'^2\frac{r_2}{f_1} - m_{14}r_2'\left(\frac{r_2}{f_1}\right)^2 - m_{13}\left(\frac{r_2}{f_1}\right)^3 + \cdots \end{aligned} \tag{5.24}$$

or equations (5.1) may be written in terms of the coefficients of the lens having the reciprocal voltage ratio

$$\begin{aligned} r_2' &= -\frac{r_1}{f_2^*} + m_{13}^*r_1'^3 + m_{14}^*r_1'^2\frac{r_1}{f_2^*} + m_{15}^*r_1'\left(\frac{r_1}{f_2^*}\right)^2 + m_{16}^*\left(\frac{r_1}{f_2^*}\right)^3 + \cdots \\ -\frac{r_2}{f_1^*} &= r_1' + m_{23}^*r_1'^3 + m_{24}^*r_1'^2\frac{r_1}{f_2^*} + m_{25}^*r_1'\left(\frac{r_1}{f_2^*}\right)^2 + m_{26}^*\left(\frac{r_1}{f_2^*}\right)^3 + \cdots \end{aligned} \tag{5.25}$$

where the asterisks (*) denote the coefficients and focal lengths of the reversed lens. A term by term comparison of equations (5.24) and (5.25) shows that each coefficient for the reversed lens is the same as one for the forward lens, with an obvious symmetry

$$m_{13}^* = m_{26} \qquad m_{14}^* = m_{25} \qquad m_{15}^* = m_{24} \qquad m_{16}^* = m_{23}$$

and also $f_1^* = -f_2$. Notice that as we have not specified which lens is the accelerating one, this set of equations can be read either way.

Using equation (5.5) we can develop the relation between the coefficients of spherical aberration:

$$C_{s_0} = -m_{13}f_2$$

so

$$C_{s_0}^* = -m_{13}^*f_2^* = -m_{26}f_2^*$$

substituting $C_{s_4} = -m_{26}(f_1^4/f_2^3)$ gives

$$\begin{aligned} C_{s_0}^* &= C_{s_4}f_2^*(f_2^3/f_1^4) \\ &= -C_{s_4}(f_2/f_1)^3 \qquad\qquad \text{as } f_2^* = -f_1 \\ &= C_{s_4}(V_n/V_1)^{3/2} \end{aligned}$$

where V_n is the potential of the last electrode and the ratio is that for the unasterisked direction. It is easy to show, by similar arguments, that $C_{s_1}^* = C_{s_3}\left(\frac{V_n}{V_1}\right)^{3/2}$

and $C^*_{s_2} = C_{s_2}\left(\frac{V_n}{V_1}\right)^{3/2}$. Knowing that the reverse magnification $M^* = 1/M$ we can write equation (5.3) for the reversed direction as

$$\begin{aligned}
C^*_{s\,(M^*)} &= \left(\frac{V_n}{V_1}\right)^{3/2}\left[C_{s_4} + C_{s_3}/M^* + C_{s_2}/M^{*2} + C_{s_1}/M^{*3} + C_{s_0}/M^{*4}\right] \\
&= \left(\frac{V_n}{V_1}\right)^{3/2}\left[C_{s_4} + MC_{s_3} + M^2C_{s_2} + M^3C_{s_1} + M^4C_{s_0}\right] \\
&= M^4\left(\frac{V_n}{V_1}\right)^{3/2}\left[C_{s_4}/M^4 + C_{s_3}/M^3 + C_{s_2}/M^2 + C_{s_1}/M + C_{s_0}\right] \\
&= M^4\left(\frac{V_n}{V_1}\right)^{3/2} C_{s\,(M)}. \qquad (5.26)
\end{aligned}$$

In the case of the afocal five cylinder lens discussed in chapter 4, the magnification is given simply by $M_{af} = (V_5/V_1)^{-1/4}$, so the ratio of the spherical aberration coefficients in the forward and reverse directions is $(V_5/V_1)^{1/2}$.

5.6.2 *Fifth order effects*

In the preceding section, we ignored the fifth order terms of equation (5.1) because, in the reverse direction, the coefficients, q^*_{ij} are not so simply related to those for the forward direction as are the third order coefficients. These can be evaluated in a similar way and are:

$$\begin{aligned}
q^*_{21} &= q_{16} + 3m^2_{16} - m_{15}m_{26} \\
q^*_{22} &= q_{15} + m_{15}(5m_{16} - m_{25}) - 2m_{14}m_{26} \\
q^*_{23} &= q_{14} + 2m_{14}(2m_{16} - m_{25}) + 2m^2_{15} - 3m_{13}m_{26} - m_{15}m_{24} \\
q^*_{24} &= q_{13}3m_{13}(m_{16} - m_{25}) + m_{14}(3m_{15} - 2m_{24}) - m_{15}m_{23} \\
q^*_{25} &= q_{12} + m_{13}(2m_{15} - 3m_{24}) + m^2_{14} - 2m_{14}m_{23} \\
q^*_{26} &= q_{11} + m_{13}(m_{14} - 3m_{23})
\end{aligned}$$

$$\begin{aligned}
q^*_{11} &= q_{26} + 3m_{16}m_{26} - m_{25}m_{26} \\
q^*_{12} &= q_{25} + 3m_{15}m_{26} + 2m_{16}m_{25} - 2m_{24}m_{26} - m^2_{25} \\
q^*_{13} &= q_{24} + 3m_{14}m_{26} + 2m_{15}m_{25} + m_{16}m_{24} - 3m_{23}m_{26} - 3m_{24}m_{25} \\
q^*_{14} &= q_{23} + 3m_{13}m_{26} + 2m_{14}m_{25} + m_{15}m_{24} - 4m_{23}m_{25} - 2m^2_{24} \\
q^*_{15} &= q_{22} + 2m_{13}m_{25} + m_{14}m_{24} - 5m_{23}m_{24} \\
q^*_{16} &= q_{21} + m_{13}m_{24} - 3m^2_{23}
\end{aligned}$$

where the m_{ij} coefficients are those for the forward direction.

5.6.2.1 Spherical aberration of the afocal lens. While the magnification of an afocal lens does not depend on the position of the object, the coefficient of spherical

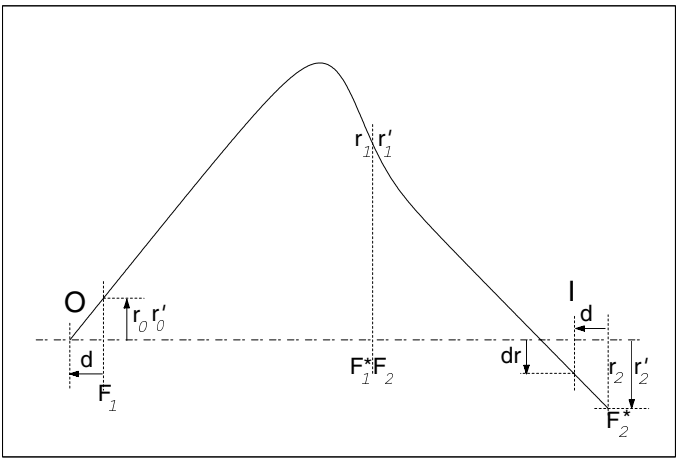

Figure 5.12 The path of a ray which leaves the axis a distance, d, to the left of the first focal plane of an afocal lens and crosses the image plane at a distance dr below the axis as a consequence of spherical aberration.

aberration does so depend. In figure 5.12 we show a schematic path for a ray which originates at a point, O, on the axis a distance, d, to the left of the first focal plane of the system. There are three focal planes to consider. The first crosses the axis at F_1, the first focal point of the first lens of the pair which constitutes the afocal combination. The second is at the second focal point, F_2, of this lens and the first focal point, F_1^*, of the second lens. The third focal plane is at the second focal point, F_2^*, of the second lens. The afocal lens produces an image of the object at a point at this same distance, d, to the left of this third focal plane.

Spherical aberration causes the ray to cross the image plane at a distance, dr, from the axis where $\mathrm{d}r = -MC_s r_0'^3$. We can use equation (5.1) to follow the path of the ray from the point where it crosses the first focal plane at a distance r_0 from the axis, to to the point at which it crosses the third focal plane at a distance r_2 from the axis. The starting point is to note that $r_0 = -dr_0'$ and then to use equation (5.1) to express the ray parameters, r_1 and r_1', at the central focal plane in terms of r_0 and r_0'. These are third order expressions. Equation (5.1) can then be used again to express the ray parameters r_2 and r_2' at the third focal plane in terms of r_1 and r_1' and so, by substitution, in terms of r_0 and r_0'. Note that all the m_{ij} coefficients are the same for the two parts of the lens. These expressions are of ninth order and an algebra package was used to calculate them. Terms above the third order were then neglected and the spherical aberration written as

$$\mathrm{d}r = -MC_s r_0'^3 = r_2 - dr_2'.$$

Substitution into this equation leads to a quite tidy expression for the aberration of the afocal lens.

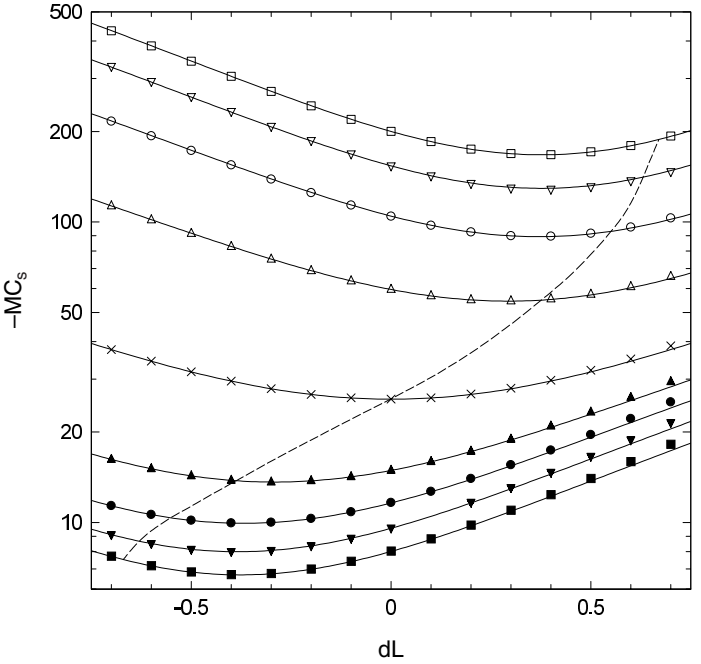

Figure 5.13 Values of the spherical aberration, $-MC_s$, for afocal lenses operated with objects a distance dL to the *right* of the symmetric position. Details of the data points and the curves are discussed in the text.

$$
\begin{aligned}
-MC_s = &\left(\frac{d}{f_2}\right)^4\left[C_{s_0} - C_{s_4}\left(\frac{f_2}{f_1}\right)^3\right] \\
&- \left(\frac{d}{f_2}\right)^3\left[C_{s_1} - C_{s_3}\left(\frac{f_2}{f_1}\right)^2\right] \\
&+ \left(\frac{d}{f_2}\right)^2\left[C_{s_2}\left(1 - \frac{f_2}{f_1}\right)\right] \\
&- \left(\frac{d}{f_2}\right)[C_{s_3} - C_{s_1}] \\
&+ C_{s_4} - C_{s_0}\left(\frac{f_1}{f_2}\right). \qquad (5.27)
\end{aligned}
$$

Figure 5.13 shows values of $-MC_s$ calculated using equation (5.27) for afocal lenses of total length 8 D and overall voltage ratios of 1, 4, 9, 16 and 25 and their reciprocals together with values obtained using the Movable Afocal routines of **LENSYS**. The abscissæ are the values of dL used in these routines and are related to the displacement, d, of equation (5.27) by $d = \mathrm{d}L - (2.0 + F_1)$. The broken curve passes through the first focal points and helps to show that the minimum

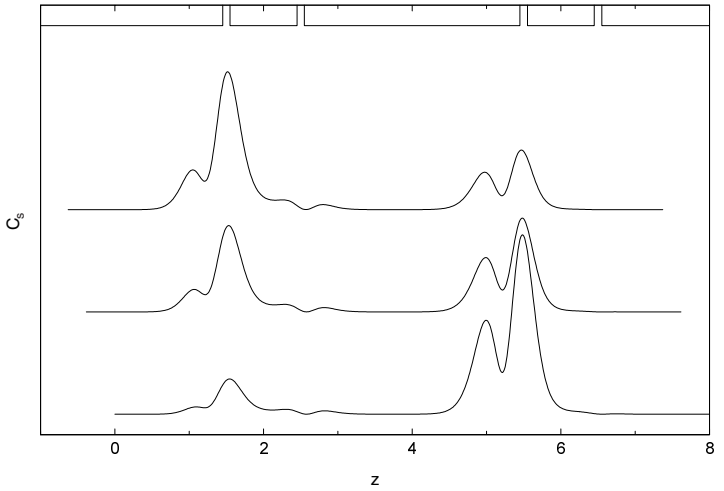

Figure 5.14 The aberration integrals for an afocal lens having a voltage ratio, $V_5/V_1 = 16$ with the object at the symmetric position, at a distance 0.38 to the left, where $-MC_s$ has a minimum value and at 0.63 to the left, at the first focal point. The curves are displaced vertically for clarity.

aberration occurs for positions between this focal point and the symmetric position. This figure shows that, for accelerating lenses, it is better to place the object to the left of the symmetric position. This has a further advantage in that, with a longer cylinder at potential V_1, it is possible to insert an angle stop, or entrance pupil, with little disturbance to the potential distribution near the first gap.

Minimum values of $-MC_s$ occur when the perturbing action is divided roughly equally between the two component lenses. This is illustrated in figure 5.14 for an afocal lens of overall voltage ratio $V_5/V_1 = 16$ with objects at the symmetric position, at the first focal point and at an intermediate position.

5.6.3 Chromatic aberration

If we consider our earlier expressions to describe accelerating lenses we can obtain results for the corresponding decelerating lenses by reversing the direction of the rays. Inspection of equation (5.22) shows that C_{c_0} and C_{c_2} must be interchanged as r_1 and r_2 are redefined so $C^*_{c_0}$ becomes identified with the term in $1/M^2$. The magnitudes of all three coefficients are changed by a factor (V_n/V_1) because $f^*_1 = f_2$ and by a further factor $(V_n/V_1)^{1/2}$ because of the $V_0^{1/2}$ term. We therefore have $C^*_{c_i} = C_{c_{(2-i)}}(V_n/V_1)^{3/2}$ where $i = 0,\ 1,\ 2$.

It is easy to show, using an analysis similar to that of equation (5.26), that

$$C^*_c(M^*) = M^2\left(\frac{V_n}{V_1}\right)^{3/2} C_c(M).$$

TSMC EX. 1016 Page 113 of 139

The position is a little more complicated for the coefficient of lateral chromatic aberration, as C_{D_1} is associated with the r_2^2 term in the integral for C_c and so $C_{D_1}^*$ must be associated with the r_1^2 in the integral for C_c^*. The magnitude ratios are just (V_n/V_1), because f_1 (f_2^*) appears only once. The result, however, is just as simple,

$$C_{D_0}^* = C_{D_0}(V_n/V_1) \qquad\qquad C_{D_1}^* = C_{D_1}(V_n/V_1)$$

though this leads to a rather strange result for $C_D^*(M^*)$

$$C_D^*(m^*) = M\left(\frac{V_n}{V_1}\right)(C_{D_1} - C_{D_0}/M).$$

5.6.4 *Figures of merit*

The geometric aberrations of a lens system for a given application may be reduced by the correct choice of the number and structure of the electrodes. The application will usually specify the potential ratio of the image and object spaces, the separation of the object and image, the magnification to be used and the maximum acceptable pencil angle of the beam in one or other of the spaces. The desirable parameter will be a minimum size of aberration disc. The diameter of the lens system might be restricted, but it is frequently a free parameter. We can write the radius of the aberration disc as

$$\frac{\Delta r}{D} = MC_{s(M)}r_1'^3$$

where we express the coefficient in units of D. The value of the diameter also appears in the actual length, L, of the system, in terms of lens parameters, as

$$\frac{L}{D} = \frac{Q-P}{D} = F_2 + q - F_1 - p = F_2 - F_1 - Mf_2 + f_1/M \tag{5.28}$$

and so we can write

$$\Delta r = \frac{LMC_{s(M)}}{F_2 - F_1 - Mf_2 + f_1/M} r_1'^3. \tag{5.29}$$

It would plainly be difficult to examine all possible lenses to find that which gave the smallest value of Δr, but, by making two reasonable approximations, it is easy to find a good criterion to point us towards a suitable lens.

The first approximation is to use equation (5.13) to express the spherical aberration coefficient in terms of the magnification and overall voltage ratio. The second is to equate the arithmetic mean of the mid-focal distances to the geometric mean of the focal lengths†. The denominator of equation (5.29) becomes

$$2(-f_1 f_2)^{1/2} - Mf_2 + f_1/M = -Mf_2(1 - 1/\sigma M)^2$$

† This result can be shown to be true for lenses weak enough that $Y = 0$, where Y is defined by equation (5.14).

and so

$$\Delta r = -LC_{s_0}(1 - 1/\sigma M)^2 r_1'^3/f_2$$

which can be separated into

$$\Delta r = -gL(1 - 1/\sigma M)^2 r_1'^3$$
$$g = C_{s_0}/f_2 = -m_{13}$$

where g contains all the lens dependent information and none of the specification and is therefore it is a suitable figure of merit for the inter-comparison of lenses. For the type of problem considered here, where the overall voltage ratio forms part of the specification, it will almost certainly be the case that a lens of three or more elements will have an appreciably smaller value of g than any two element lens. Once the lens geometry and inner voltage ratios have been selected, the choice of lens diameter follows from equation (5.28) where the focal properties are in units of D.

For the case of rays focused from infinity, for which the magnification is zero, this analysis fails because the distance between conjugates is infinite. However, the appropriate specification in this case is more likely to be in terms of the distance of the image from the lens itself, probably specified as $L = F_2$, the mid-focal distance. The radius of the aberration disc is given by equation (5.6) and the figure of merit is just $C_{s_4}/f_1^3 = -m_{26}(f_1/f_2)$, though, as the ratio of focal lengths is implicit in the specification of the voltage ratio, $-m_{26}$ is itself suitable.

If the specification allows for an adjustable magnification, the quantity to be optimized becomes the product MC_s, which, again using equation (5.13) can be written as

$$MC_{s(M)} = MC_{s_0}(1 - 1/\sigma M)^4.$$

We may differentiate this to find an optimum value of $M = -3/\sigma$, which gives for the minimum radius spot in the Gaussian image plane

$$\Delta r = -9.48(C_{s_0}/\sigma)r_1'^3 \quad = 2.37C_{s_1}r_1'^3$$

and suggests that m_{23} would be a better figure of merit with this looser constraint.

6 The LENSYS program

6.1 The programs on the disk

The disk which accompanies this book contains a number of files for use on an IBM or compatible personal computer. There are two versions of the **LENSYS** program, one, **LENSYS2.EXE**, compiled in real mode and the other, **LENSYSP.EXE**, in protected mode. The difference between the two compilations is that **LENSYS2** runs in the lowest 640kB memory of the computer and is therefore suitable for a basic PC, while **LENSYSP** will use memory above this region and will normally be the version of choice. The program source is exactly the same for the two compilations.
The files are arranged in three subdirectories:

PROTECT	REALMODE	RELAX
DPMI16BI.OVL	LENSYS2.EXE	RELAX51A.EXE
LENSYSP.EXE	LENSYS2.ICO	RELAX51A.ICO
LENSYSP.ICO	LPOTS.DAT	RELAX51P.EXE
LPOTS.DAT		RELAX51P.ICO
RTM.EXE		RELAXCYL.EXE
		RELAXCYL.ICO

The file **LPOTS.DAT**, which is loaded automatically by **LENSYS**, contains data from which the potentials and potential gradients are calculated. The files **DPMI16BI.OVL** and **RTM.EXE** are required to run any protected mode program. The **RELAX** subdirectory contains the auxiliary programs which are discussed in section 6.7.

The programs have been written to run under MSDOS, and will run as DOS programs under Windows 98. The various icon (**.ICO**) files are provided so that Windows shortcuts may be generated for all the programs and placed, if desired, on the desktop. **LENSYS2.EXE** itself requires some 400kB of memory and, if this is not available, the program will stop with a message that there has been a memory overflow error. This might occur on a basic PC if several TSR programs have been installed, but is unlikely to occur on a modern machine. If memory is not a limitation, then **LENSYSP.EXE** is the better choice.

A graphics screen is necessary: VGA is best, but EGA is acceptable and, though the program will use a CGA screen, the resolution is only just adequate, especially for text messages and data tables. The Hercules monochrome graphics screen is usually very clear, but the lack of colour is not a trivial loss. The one place where the program behaves differently on a monochrome screen is in the menus, in which the selected item appears in inverse video rather than in a different colour. All the graphics routines are specific to the programs and do not use any of the facilities of Windows.

6.1.1 Loading and running the programs

None of the programs should be run from the floppy disk, because any data file they may write is put into the directory from where they are run. All the files in the **REALMODE** or **PROTECTED** subdirectory should be copied to a dedicated subdirectory on the computer hard disk. **LENSYS** will only look in its own directory for the **LPOTS.DAT** data file and the protected mode support files

In DOS mode the program is started by typing the filename at the DOS prompt. In Windows mode it can be started using the RUN command, double clicking on the filename in Windows Explorer or on the shortcut icon if the user has prepared one. It presents a screen with brief details of the program. If it does not find a maths coprocessor in the computer, it writes a message to the screen telling you that it would run much faster with one, but it continues nevertheless. Once this introductory screen has been displayed, it reads the data file and will then make no further use of the program on the disk. After the data file has been loaded, you are invited to 'Press **ENTER** to continue'. This starts the program and shows a screen which will provide the communication channel.

6.2 The screen display

The lower three quarters of the screen is the graphics window. This will show schematic representations of the lens being studied and the rays traced during the analysis of the lens. The breadth of this window is equivalent to eight lens diameters. The upper portion of the screen forms two text windows. The wider one, to the right, is called the data window and is for the display of lens data and the results of the calculations, while that to the left is called the message window and shows the menus, together with other prompts and messages in the course of the program. It is usually of a different colour from the data window. The manual input of lens parameters (potentials, lengths etc) is normally made to the data window, but sometimes to the message window.

6.2.1 The menu system

All choices are made via a system of menus, each offering between two and five options. The opening menu describes itself as 'Main Menu' and invites a choice between

Set Options
Lens Data
Imaging
Exit to DOS

with the top entry displayed in a different colour (inverse video on a monochrome screen) to signify that it is the 'selected' item. The selection may be changed by the UP- and DOWN-arrow keys and, as the motion rolls over, the quickest way to select the bottom item is to use the UP key once. Pressing the ENTER key makes the choice. Choosing 'Exit to DOS' is the normal exit route from the program. If it is chosen, the program does not request verification, but exits immediately with a simple closing message leaving the display in a 25 line text mode. The program does not remain in memory and has to be reloaded if required again.

6.3 The Set Options option

Choosing 'Set Options' takes you to an 'Option Menu' which offers the choice of

Show V-axial
Show Cs
Show both
Show neither
Save all

This choice enables the display of the selected data in addition to the ray diagrams. After making the choice, you are returned to the main menu. The axial potential and the integrands of the spherical aberration integrals will be written to files if they are selected to be shown. The 'Save all' option saves, in addition to these, the integrands of the chromatic aberration integrals and the Petzval integral, and, though these are not displayed, they can be examined with a suitable graphics program.

6.4 The Lens Data option

Choosing 'Lens Data' takes you to another menu 'Select a Lens' which offers the choice of

2-Aperture A=D/2
2-Aperture A=D
3-Aperture A=D/2
2-Cylinder G=D/10
3-Cylinder G=D/10

This section allows you to find the focal properties of a lens. The geometries of the first four lenses are predefined, but if you choose the three cylinder lens you are asked, in the message window, for the length of the second element, in

units of the diameter, and this should not exceed 3. If you enter too large a value, the invitation is repeated. No sign is expected, three places of decimals will be accepted, but the length will be rounded down to the next multiple of 0.0125. The program then draws a schematic of the lens in the graphics window and asks for the voltage ratios, V_2/V_1 and V_3/V_1 (if appropriate). These are entered in the message window. Up to three decimal places are accepted here. All voltage values presented in the data window are on the basis that $V_1 = 1$ and so V_n represents V_n/V_1. The chosen values of the electrode potentials are written in the graphics window, above the schematic electrodes. The message window will tell you that the program is calculating the trajectories and the aberrations of the lens† and two rays will be drawn in the graphics window. These are the usual principal rays, drawn from the left and from the right, parallel to the axis. The focal points are marked and labelled as are the principal points. The axial potential and the five spherical aberration integrands will also be shown if they were selected from the Options Menu. The data window shows three columns of data. The first column displays values of the focal lengths, mid-focal distances and the positions of the principal points, the second column values of the five spherical aberration coefficients, C_{s_i} and the Petzval sum, and the last column shows values of the five chromatic aberration coefficients, C_{c_i} and C_{D_1}.

When the calculation and display are complete, the message window shows a further menu headed 'Continue?'

Swap Cs/Mij coeff
New Potentials
New Lens
Print/Save
Main Menu

We shall discuss 'Print/Save' later. The first choice, 'Swap Cs/Mij coeff' changes the display in the right hand section of the data window to show the eight m_{ij} coefficients and the four elements of the bending matrix described in chapter 1. A repeated choice restores the original display. The effect of choosing 'New Potentials' is obvious. The 'New Lens' option leaves you in the 'Lens Data' part of the program by returning you to 'Select a Lens'. If you wish to examine the properties of a lens between conjugates it is necessary to choose 'Main Menu' and then 'Imaging' from that menu.

6.4.1 Error conditions

There are a number of input parameter values which the program will not reject immediately, but it will discover that they are not sensible and will halt with an appropriate message and wait for ENTER to be pressed. A very simple error is to input $V_2 = 1$ for a two element lens which gives no lens action at all. Another input leading to error conditions is an excessively high voltage ratio. There are a number of tests for these. If the emergent ray has the wrong *sign* of slope it has

† The speed of a modern PC is such that these messages may not be seen.

probably crossed the axis twice. If the ray crosses the axis between the outermost electrodes, the focal length will not be a very useful parameter and it would be misleading to display the results in the data window.

It is possible to input a zero or negative value for a potential. This would be nonsense in the case of the final electrode and the program will reject the negative sign, but it may be acceptable for the centre electrode of the three aperture lens or a cylinder lens having a very short centre element. As an example, the three aperture lens having $V_3 = 1$ will accept values down to -0.8 for V_2. A zero value gives the condition for the classic Einzel lens.

6.5 The Imaging option

This opens with a further menu, 'Select Lens Type', and you have to choose between

Aperture lenses
Cylinder lenses
Special lenses

The details of the lens specifications and the positions of the conjugate points have to be entered in a slightly different fashion for the two lens types. These differences will be explained later. The program traces a ray starting from the axial object point at the left of the graphics window and with a slope of 1, but the subsequent action depends on a further choice which you be asked to make by a menu headed 'Solve the Lens'.

Adjust Potentials
Find the Image

6.5.1 Types of solution

6.5.1.1 Automatic adjustment. If you choose the automatic adjustment of a potential, the potential of one electrode will be adjusted, in a series of steps, until the ray crosses the axis at the image point you have specified. Each ray is deleted as the next ray is being drawn. In the case of lenses with two or three elements the potential of the second electrode is adjusted, but if there are more than three electrodes you are asked, by a message, to indicate which one. You may not choose the first or the last electrode. After the program has made an automatic adjustment it draws two further rays. These both start from an object point above the axis and one has a slope of zero and the other a slope of -1. These rays will cross at the image plane and indicate the magnification. The data window shows the lens specification (a little differently in the various cases), the electrode potentials, updated at each iteration step, and the magnification and the spherical and chromatic aberration coefficients.

For any lens geometry, more than one set of electrode potentials will satisfy the conjugate focus condition. It is possible to persuade the program to pick one combination rather than another by choosing a suitable starting point for the potential which is to be varied. For example, even a two element lens has two possible voltage ratios, one accelerating and the other decelerating, and if V_2 is initially set to 0.1 the program should find the retarding solution. However, if this requires a potential of less than 0.01 it will fail, with a message 'This is unlikely to converge. ENTER continues'. A three element lens started with $V_3 = V_2 = V_1 = 1$ will converge to the solution with $V_2 > 1$. There is a general tendency built into the program to look for the higher potential solutions because these are usually the ones with smaller aberrations.

6.5.1.2 Error conditions. If the automatic routine cannot find a value which allows the ray to pass through the axial image point, it will issue a message to that effect and wait for ENTER to be pressed. This will sometimes happen because it is looking in the wrong region. For example, it might be trying very small values and meeting the $\not< 0.01$ constraint while a much larger value may be valid. This condition can usually be caught by noting the way in which the variable potential changes after each iteration. This may not be possible with the speeds at which modern PCs operate, though there are ways, discussed in the appendix, by which this factor may be circumvented.

A more common cause for failure is that the overall voltage ratio is too high. If the potential of the variable electrode is observed to be varying around, or near to, the geometric mean of the neighbouring potentials, this can be taken as a sign that that region of the lens is too strong and new fixed potentials should be tried. It is often worthwhile to try some values close to these critical ones, using the 'Find the Image' option, to assess the nature of the change that is required.

6.5.1.3 Find the image. If you choose to find the image, only a single ray is drawn and the program will find the position of the image point appropriate to the potentials you have chosen and will draw a vertical line through that point. The image may lie beyond the right hand end of the screen, but the position is noted in the data window. The aberration coefficients and the magnification are calculated and shown in the data window.

You are then asked if you wish to draw another ray. If so, you are asked for the starting position and slope of the ray and it is then drawn and the 'Another Ray?' question repeated. None of these rays are deleted until you exit by declining the invitation. This mode allows a particular set of rays to be drawn for a lens for which the automatic adjustment has already provided values of the electrode potentials. It might, for example, be used to find the position of the image of a second aperture, an entrance pupil say, which can be simulated by the intersection of a ray parallel to the axis and some distance above it with the original ray.

In this mode saving the data files will only save the first, image finding ray. Other rays can only be recorded by a screen print.

At the end of the lens solution routines the message window shows the same 'Continue?' menu as in the Lens Data case, but now the choice of 'New Lens' returns you to the 'Select Lens Type' menu.

6.5.2 Aperture lenses

A menu offers you the choice of the three aperture lenses already seen in the Lens Data option. A message asks for values of the object and image distances, P and Q, with a constraint that the total distance from object to image should not exceed eight diameters. Despite the sign convention requiring P to be negative, only the magnitude should be entered as the program can only handle real objects and images and there can, therefore, be no ambiguity. The values appear in the data window. A schematic diagram of the lens is drawn in the graphics window, with the object at the left hand edge of the window, and the centre of the lens a distance P to the right. A vertical line is drawn across the axis a further distance Q beyond the lens centre. You are then asked for the electrode potentials which are written in the data window replacing '?' prompts. Next the 'Solve the Lens' menu appears and the subsequent behaviour is as we have already discussed.

6.5.3 Cylinder lenses

A menu, 'Select a Lens', offers a choice of various cylinder lenses

2-5 Cylinder
Afocal6
Afocal8
Varimag6
Varimag8

The last four of these will be discussed later. Choosing the first option leads to yet another menu, 'How many cylinders', and thence to the particular multi-cylinder lens you require.

In contrast to the aperture lenses, for which only a limited set of spacings are available in this program, the cylinder lenses may use a range of lengths of cylinder and it is rather simpler to specify the object and image positions in terms of these lengths. This is particularly the case for a lens with cylinder lengths of (say) 2, 1, 2 and 3 diameters where the choice of reference plane might be a matter for argument. The program, therefore, expects the object to be at the left hand end of the first cylinder and the image at the right hand end of the last cylinder, though it will extend the last cylinder to the right hand edge of the screen in the schematic diagram, marking the image plane with a transverse line as before. Having chosen the number of cylinders, you are asked for their lengths (which will be rounded down to the next multiple of $D/80$) and their potentials, which may be specified to three decimal places. The values of the lengths are taken to include a half share of the inter-cylinder gap of $D/10$ and the total length must not exceed eight diameters. As a trivial example, a lens with $L_1 = 2.00$, $L_2 = 1.00$ and $L_3 = 3.00$, for which the reference plane is unambiguously in the middle of the second element, would

have $P = 2.50$ and $Q = 3.50$. The 'Solve the Lens' menu then appears and the subsequent behaviour is as we have already discussed. Lenses with three cylinders are widely used to satisfy the simple condition of maintaining a conjugate pair of points for a range of overall voltage ratios. An example was shown in figure 4.6.

The length of the centre element affects the behaviour of both the magnification and the spherical aberration, though not to a great degree. The addition of a fourth cylinder allows for better control of the magnification, but a fifth cylinder opens the possibility of controlling not just the magnification, but also the spherical aberration of the lens system.

6.5.3.1 The aberration behaviour of the five cylinder lens. The conjugate focus condition for a lens having the potentials of the third and fifth electrodes fixed may be achieved with each of the intermediate potentials at a high or low value compared with their neighbours. The four possible combinations lead to significantly different coefficients of spherical aberration for all values of the magnification. This is illustrated in figure 6.1 for a five cylinder lens having $V_5 = V_3 = V_1 = 1$. For all values of the magnification, the spherical aberration coefficient is least for $V_2/V_1 > 1$ and $V_4/V_3 > 1$.

6.5.4 The Afocal and Varimag lenses

These lenses each have five cylinder elements with lengths

for the "6"	and the "8" lenses
$L_1 = 1.25$	1.50
$L_2 = 0.50$	1.00
$L_3 = 2.50$	3.00
$L_4 = 0.50$	1.00
$L_5 = 1.25$	1.50.

6.5.4.1 The Afocal lens. The afocal lens was discussed in chapter 4 and the only input required is the value of the overall voltage ratio. The program automatically sets $V_3 = V_5^{1/2}$ and adjusts the other potentials maintaining $V_4/V_3 = V_2/V_1$. The data window reflects the changes in the potentials and, after convergence, shows the magnification and the spherical aberration coefficient. The program operates only in the regime with V_2 and V_4 both at the higher of their two possible values. The shorter lens has much smaller aberrations, but the end elements are rather short. Placing a real angle stop within the electrode would therefore affect the potential distribution to a certain extent.

The behaviour of a non-symmetric afocal lens of total length 8 D is discussed later in section 6.5.5.1.

6.5.4.2 The Varimag lens. These are very powerful lenses which allow the magnification and the overall voltage ratio to be specified and a further parameter

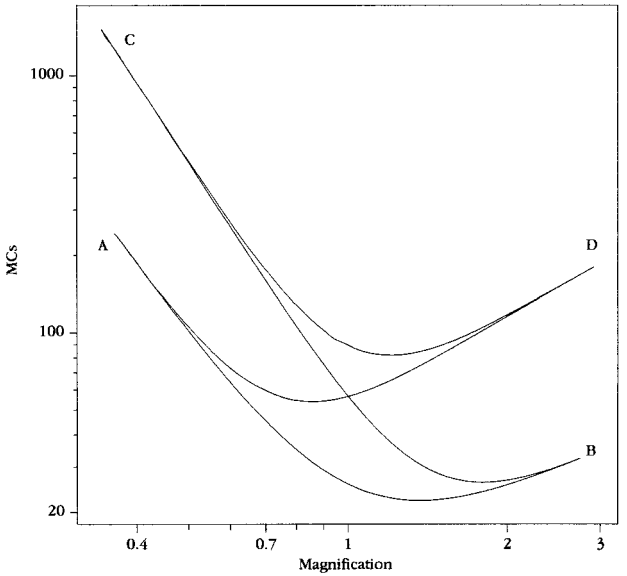

Figure 6.1 The spherical aberration, expressed as the product MC_s, of a five cylinder lens with $V_5 = V_3 = V_1$ as a function of the magnification for values of V_2 and V_4 in the following ranges:

Region	V_2	V_4
AB	> 1	> 1
BC	> 1	< 1
CD	< 1	< 1
DA	< 1	> 1.

to be adjusted. This may be either the magnification of an entrance pupil or the spherical aberration coefficient of the lens. These are controlled, though not independently, by the potential of the centre electrode. The program assumes that there is an entrance pupil at a distance $D/2$ from the entrance window. You are asked to supply values for V_5, V_3 and the target magnification. This will be a negative quantity, but your input will be read and displayed in the initial data as the modulus, |MagT|. The program first examines the range of magnifications which can be achieved with these potential values, and may halt with a message to the effect that the target magnification is outside the usable range. Otherwise it adjusts V_2 to satisfy the conjugate condition and then adjusts V_4 to change the magnification. Both V_2 and V_4 are constrained to take the higher of their possible values.

These adjustments of V_2 and V_4 continue alternately until the window magnification is close enough to the target value. At this point the data window,

which has monitored the potentials and the window magnification throughout, displays the target, window, pupil and longitudinal magnifications (MagT, MagW, MagP & MagL) and the spherical aberration coefficient of the lens. A menu invites a choice of

Modify V3
Continue

The first allows you to change V_3 and the whole convergence procedure is repeated while the second opens the usual 'Continue?' menu.

Increasing V_3 from the 'afocal' value of $V_5^{1/2}$ will both decrease the pupil magnification and reduce the spherical aberration. If the target magnification is close to one of the limiting values with the initial value of V_3, it is possible that changing V_3 might modify the limit sufficiently to cause the program to fail to converge. If this happens, it will be necessary either to change the target magnification or to accept a restriction on the value of V_3.

In the case of the six diameter lens, the exit pupil can be seen, provided that the longitudinal magnification is less than four. The reader will notice, if this lens is operated with $V_5 = V_3 = 1$ and MagT$= -1$, that the pupil magnification is not quite equal to -1. This is because the potential 1.25 diameters inside a cylinder differs sufficiently from the potential of the electrode itself that the trajectories are not quite straight. The effect is significant only at the 1% level and, for the eight diameter lens, it is even smaller.

6.5.5 The special lenses

The 'Special Lens' menu has two choices:

Movable Afocal
Excitation Lens

6.5.5.1 The movable afocal lens. This is the same as the 'Afocal8' lens of section 5.4.1, but with provision for moving the object point along the axis by an amount dL, which may be positive or negative, but should not exceed 1 D. You are asked for a value for dL and for the overall voltage ratio, V_5. The program behaves exactly as the 'Afocal8' program and then offers a choice

Change dL
Continue

'Continue' takes you to the 'New Potentials' etc menu and that choice returns you to the lens, with the previously chosen value of dL, and allows you to set a new value for V_5. This procedure is slightly more clumsy than for the preset geometry of 'Afocal8', but the two variables allow you to study the dependence of C_s on both parameters. While the magnification of an afocal lens depends only on the overall voltage ratio, the spherical aberration depends also on the location of the object and image points within the lens. This behaviour was discussed in chapter 5.

All the data for this lens are written to a file **MOVFOCAL.DAT**. Note that this file may be added to if 'Movable Afocal' is chosen again during a single run of **LENSYS**, but will be overwritten if the program is stopped and restarted.

6.5.5.2 The excitation lens. This is a five cylinder lens developed for one purpose, to study the excitation of a gas target by a beam of electrons and the subsequent radiative decay. The requirement is for a reasonably constant potential in the observation region at the centre of the lens and some form of potential barrier to restrict secondary electrons, formed at beam-forming apertures, from entering this region. It is geometrically and electrically symmetric with V_2 and V_4 operated, usually, at a lower potential than V_3. An application of the lens to a specific problem is discussed by Heddle and Kay [18].

You are asked for values of the potentials and to specify which one should be varied. After solving the lens, the program draws two rays of slope $\pm$ 1 from an axial object point and displays the magnification, the spherical aberration coefficient and the radius of the trajectory at the centre of the lens, r_{mid}, in units of the lens diameter.

6.6 The Print/Save option

The lens data, the ray paths and the axial potential and spherical aberration integrands if selected, may be written to files. A simple pixel level dump of the screen can be made to a suitable printer connected to the first parallel port (LPT1) of the computer.

On selecting Print/Save from the 'Continue?' menu, a second menu, 'Printer Type' invites a choice of

Save to Files
'Laser'
'Matrix'

The terms 'Laser' and 'Matrix' are meant to define respectively a 300 dpi laser or inkjet printer and a nine pin dot matrix printer.

6.6.1 The Save to Files option

Choosing 'Save to Files' will cause a counter to advance by one and the data for the lens distances, potentials, aberration coefficients and magnifications to be written to a file **LENSnnn.DAT** where **nnn** is the number of the lens. This is reset to zero when **LENSYS** is restarted and incremented when each lens for which 'Save to Files' is selected. The ray paths are saved to a similar file with extension **.RAY**. If a choice has previously been made to show the axial potential or the aberration integrand, then these will be written to files with extensions **.POT** and **.ABR**. If the 'Save all' choice has been made, then the integrands of the chromatic aberration integrals will be written to a **.CHR** file and, in the 'Lens data' mode, the integrand of the Petzval integral will be written to the **.ABR** file.

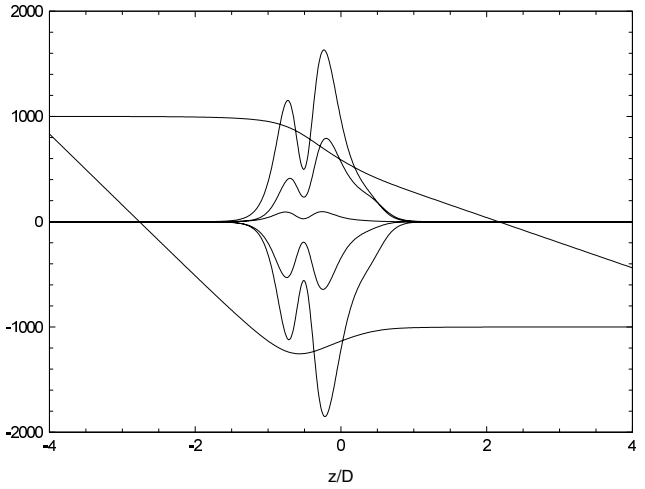

Figure 6.2 The ray and aberration data for a two aperture lens plotted using the **.RAY** and **.ABR** data saved from the same run as the screen dump of figure 5.11.

If the choice of 'Find the Image' ray tracing has been made the **.DAT** file shows the image position. Even if a second ray is traced, only the original ray will be saved to the **.RAY** file, because further rays would overwrite the first one. It was not practicable to use different ray variables as there is no limit to the number of additional rays, for various launch positions and angles, which may be traced.

The saved **.RAY, .ABR, .CHR** and **.POT** files can be displayed and scaled by a suitable graphics program and show much smoother variation than the screen copies. Figure 6.2 shows the same data as figure 5.11 displayed using the 'EasyPlot' program†.

All the data files are properly normalized and numerical integration of the saved data will give the same values as listed in the **.DAT** file.

A complete save of the data files for a 'lens data' run amounts to some 125 kB and for an 'imaging' run to some 70 kB. On leaving the program, the exit screen will remind you if you have saved any files. You should examine these before you restart the program as they will be overwritten, or set to zero by fresh saves.

6.6.2 The printing options

The other choices from the 'Printer Type' menu lead to a 'Which Region' menu

† Cherwell Scientific Publishing Limited.

Whole Screen
Data Window
Graphics Window
Continue

The 'Continue' option is offered in this menu, because a choice of 'Save to Files' for a lens for which this is not possible, would otherwise force you to an unwanted 'Print'. This gives a safe exit.

The message window then clears and some information about the lens is displayed there. The printing is done by reading each pixel and copying it as a black dot unless it has the colour of the background. The DOS Graphics command is not used and the quality of the printed output reflects that of the monitor, but with all the colours lost. Figures 2.1(b), 5.10 and 5.11 were printed from a VGA monitor. The program issues two Line Feed characters after printing, but no Form Feed character. A VGA display printed to a laser printer retains the correct proportions and is of such a size that two complete screens or seven data windows may be printed on A4 paper.

If the printer is not connected, not turned on or has run out of paper a message will be issued and the program will wait for ENTER to be pressed, at which point it will return to the 'Continue?' menu.

This method of obtaining a printed image of the screen will only work if the printer can interpret the codes of the PCL level III language, in the case of a laser/inkjet printer, or of the Epson FX language in the case of a dot matrix printer. Many modern laser/inkjet printers cannot do this as they rely on drivers built in to the Windows system.

6.6.3 Printing under Windows

If the program is run under Windows, a copy of the screen can be saved to the clipboard by pressing the **PrintScreen** key and can be pasted into the **PAINT** program using **Ctrl V**. It can then be printed in colour to any Windows printer. This is not very satisfactory, because the blue background, which is very suitable for visual display, does not offer adequate contrast when printed. More sophisticated painting programs are available which allow the colours to be changed and also allow the screen to be captured and cropped. A white background is the most effective for printing, but any white areas, such as text, should first be changed to (e.g.) black.

6.7 The auxiliary programs

These are included to illustrate the manner in which the potential distribution approaches its final value. They also allow the oscillatory behaviour, which occurs with large values of g, to be observed.

6.7.1 The **RELAX51A.EXE** *and* **RELAX51P.EXE** *programs*

These programs illustrate the solution of the two dimensional problem described

in section 3.4.1 using the five point relaxation methods. The initial condition is that the potential along the left hand side and the top of the 51×51 array is set to 1, while the rest of the array is set to zero except for the bottom left and top right corners which are set to 0.5.

They differ in the order of scanning the array of points. **RELAX51A** follows the sequence

```
for x:=1 to 49 do
for y:=1 to 49 do
```

while **RELAX51P** follows the sequence

```
for x:=1 to 49 do
for y:=49 downto 1 do
```

The effect of these sequences is to drive the potential distribution along the $x = y$ diagonal and perpendicular to that diagonal, respectively, which gives a more symmetrical distribution along the diagonal at all stages.

In both cases the criterion for satisfactory completion of the iterative procedures is more severe than that illustrated in figure 3.8. Instead of just requiring that the sum of the potentials on the diagonal should be smaller in magnitude than some value, these programs require the sum of the absolute differences between the individual diagonal potentials and 0.5 to be below some level. In the original program, it would be quite possible to satisfy the criterion while having a systematic excess for $x(= y) < 25$ and a systematic deficit for $x(= y) > 25$. The different scanning modes of these two programs allow the asymmetry in the axial potential to be seen to develop in the 'A' version, while a symmetry is imposed by the 'P' scanning mode.

On running these programs you are asked for a value for the acceleration parameter to be used, and the value of the sum of the potentials of the points on the diagonal is displayed on the screen after each cycle of iteration. This sum is also saved to a file identified by the value of the acceleration parameter.

The screen also displays an array of 51×51 small squares which are coloured to represent the potentials at each point. The way in which the potentials change with successive iterations can be seen and the effect of acceleration and, especially, excessive acceleration is very clear.

6.7.2 The **RELAXCYL.EXE** *program*

This treats a similar problem in cylindrical coordinates, the two cylinder lens having a narrow gap. The five point relaxation method is again used here and the criterion for satisfactory completion is one which is appropriate for a demonstration, but is not as safe as that used for accurate calculations. The change in the potential at each point on the axis after each stage of iteration is determined and these values are added algebraically to give a figure, ΔV, and then their moduli are added to give a second figure, $|\Delta V|$, which is never negative. This latter value decreases, though not without limit, and the iteration is stopped when it becomes less than some

value, which, in this example, is set as 0.000 01. Accurate methods of assessing convergence are described in the appendix. The method used here corresponds to the convergence of the axial potential to a few parts in 10^{-5}, which is quite adequate for the very coarse display of the potential used in this demonstration.

The lower part of the screen displays the iteration number, ΔV and $|\Delta V|$. Notice that these last are equal for small values of the acceleration parameter, but ΔV changes sign to a greater or lesser degree as this parameter is increased.

The screen displays a representation of the electrode system with the potentials at each point represented by coloured squares. In contrast to the square array problem, potentials in adjacent regions differ by constant ratios, i.e. the colours are chosen to represent the *logarithm* of the potential. The effect of too large a value of the acceleration parameter is quite dramatic.

Appendix: Technical aspects of LENSYS

A.1 The programming language

The **LENSYS2.EXE** program is written mainly in Borland Pascal with some sections in Assembly Language. The floating point variables are of type double and require a maths coprocessor or its emulation: this latter is included in the code and will be used automatically if needed.

A.2 Initialization

The program first looks for a maths coprocessor and, if it does not find one, writes a message to the opening screen. It then loads the data file, stores the contents in dynamic memory and waits for a response from the user. The type of graphics adaptor is then examined and various scale and colour parameters are set in the light of this. If a CGA adaptor is found, it is set to the four colour resolution of 320 by 200. This does not affect the effective resolution of the display seriously and is preferred to the two colour 640 by 200 mode.

A.2.1 The effect of processor type on the speed of the program

When the program was first written, some ten years ago, typical processor clock speeds were 10–25 MHz and the type of processor was significant. The clock speeds in modern personal computers are some 20 times as fast and even the longest calculations, such as the Varimag routines, are completed in under a second. This is not an unalloyed gain, as there are occasions when it would be of interest to observe the convergence process. There are various programs available which can slow the effective clock speed† and so allow the successive traces to be observed.

† For example, SLOWDOWN.COM, which can be downloaded from members.aol.com/bretjohn.

TSMC EX. 1016 Page 131 of 139

A.3 Calculation of potentials

The file **LPOTS.DAT** contains values of the axial potential and potential gradient for all the lens geometries at intervals of $D/80$, which is the step size used by the ray tracing routines, from the centre to a distance 8 D into the lens. The potentials were calculated at closer intervals to improve both the precision and the accuracy of the data.

A.3.1 Aperture lenses

The potential distributions for the aperture lenses were calculated by the nine point relaxation method of equation (3.23) using, in the notation of figure 3.10, $D_1 = 160$, $D_2 = 480$ and $D_3 = 800$ with a length from the centre of the lens to one end cap of 2400 and half-gaps, $A/2$, of 40 or 80. By using a length of 5 D_2 the effect of the end cap is completely unimportant in determining the potential at 8 D_1 from the lens centre.

A.3.2 Cylinder lenses

The potential inside two closely spaced ($D/10$) cylinders is very well represented by an expansion in terms of Bessel functions and the particular method described in Section 3.6 was used in the original version of **LENSYS**. This required an estimate of the gap length correction. This correction is now known to a much greater accuracy, but only as a result of comparison with the very accurate results of the nine point relaxation method [19] and so we have used these results directly in the present version. The diameter of the cylinder was 320, the length, from centre to end cap, 3200, and the gap half-width 16 all in units of the lattice spacing.

A.3.3 Details of the calculations

The space within the half-lens was represented by two arrays. The main, cylinder, array had dimensions 241 × 2401 for the aperture lenses and 321 × 3201 for the cylinder lens. The gap array had dimensions 160 × 41 or 81 for the aperture lenses and 160 × 17 for the cylinder lens. The relaxation procedure is started, using equation (3.23*a*) at the outermost point at the midplane and progresses row by row down to the cylinder radius when the pattern changes to column by column advancing axially to the closed end of the electrode. The axial potential is then calculated using equation (3.23*b*).

Though the lens has a nominal gap of $D/10$, the finite spacing of the lattice of points on which the potential is calculated leads to an *effective* gap of 0.09964, some 0.3% less. This not a serious discrepancy.

The convergence of the iteration is assessed by an automated version of that illustrated in figure 3.14. Values of the axial potential are recorded after some number, n, cycles of iteration and are compared with values found after a further δn cycles. The initial value of n is usually 1000 and that of δn 100. The test is whether $|((V_n(z,0) - V_{n+\delta n}(z,0))/V_{n+\delta n}(z,0))| < 10^{-14}$ as z is advanced from the midplane. At the point where this test fails, the values of $n + \delta n$ and $(z-1)/D$

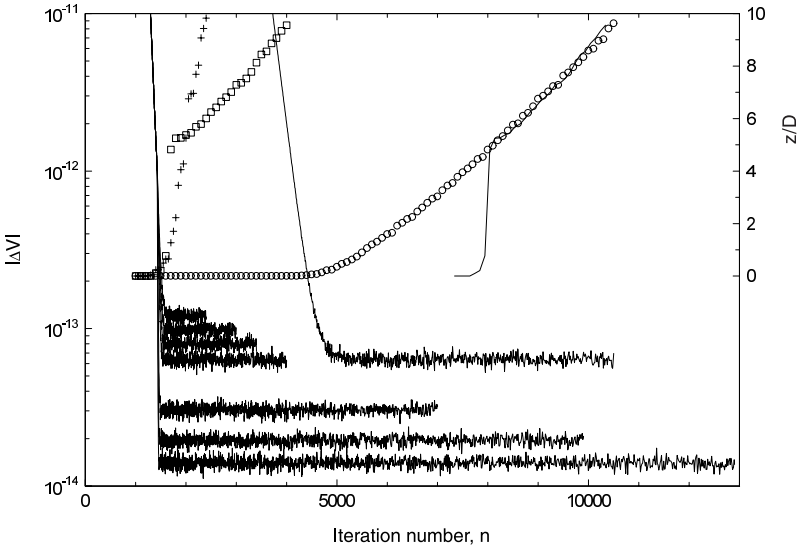

Figure A.1 The convergence of the iterative process. The left hand ordinate shows values of $|\Delta V|$ for iterations started with $g = 1.980$ and, when $|\Delta V| < 10^{-12}$, continued with values of (from the top) 1.973, 1.967, 1.960, 1.950, 1.900, 1.850, 1.800. The result for iteration with $g = 1.950$ throughout is also shown. The right hand ordinate shows the approach to convergence for the fastest choice of g : 1.980 initially and 1.973 later (+), and for $g = 1.950$ in both modes. The full line interwoven with the ○ data is discussed in the text.

are recorded, n is reset to $n + \delta n$ and the iteration is continued until the test is satisfied out to $z/D > 9.5$. There is no point in testing further as the potential is affected by the end cap and, in the present application, values greater than $z/D = 8$ will not be used.

Figure 3.9 suggests that an acceleration parameter, g, somewhat greater than 1.97 should give optimum convergence and the results from a sequence of values show that $g = 1.972$ is, indeed, best giving convergence after 4700 cycles of iteration, though the minimum is not very sharp.

This value was obtained with a single acceleration parameter, but we have seen earlier (figure 3.11, for example) that the initial approach to convergence may be speeded up by using a significantly greater value until the magnitude of the changes in the potential have decreased towards their final value. We show in figure A.1 the behaviour of $|\Delta V|$ obtained using $g = 1.98$ until $|\Delta V|$ has decreased to 10^{-12} and then smaller values. The different asymptotic values of $|\Delta V|$ are clearly shown, covering a range of almost 1 : 10 for $g = 1..800 - 1.973$ which latter value gives convergence after 2400 cycles. The progression of this convergence is shown by the + symbols. Increasing the acceleration parameter to 1.974 increases the number of cycles for convergence to 4900. This would overwrite the data for $g = 1.973$ and is, therefore, not shown.

The figure also shows, by □ symbols, the convergence for $g = 1.950$ after the initial over-acceleration. The interesting feature here is that convergence to $z/D \approx 5.5$ occurs very rapidly. This is always the case if the second value of g is significantly smaller than the first. Even if $g = 1.00$ there is still this very rapid convergence to $z/D \approx 5$. Notice that this is faster than the initial convergence for $g = 1.973$ which gives the fastest *complete* convergence. This *step*, for that is how it appears with $\delta n = 100$, is associated with the even more rapid fall in $|\Delta V|$ when the g-value changes. This very rapid convergence, assessed as it is by the values of the *axial* potential, does not mean that the potential away from the axis has reached its final value. A rough guide to the behaviour with excessive acceleration can be found by running the **RELAXCYL** program. The advancing equipotentials are clearly not of the same shape as for more reasonable acceleration, though the final state does represent a complete convergence. It can be helpful to PAUSE the program if the computer speed is too high.

The fall in $|\Delta V|$ obtained with a single value, 1.950, of the acceleration parameter, and the associated approach to convergence (○) are also shown. The full line which weaves through these points is a copy of the □ data displaced to the right by 6340 and shows that, after the initial rapid rise, the behaviour is controlled by the final value of g. It is very similar to the $g = 1.973$ behaviour (+), but with the abscissa stretched by a factor of seven.

The optimum number of cycles of iteration is determined by a balance between the convergence criterion and the limiting value of $|\Delta V|$. This is *roughly* inversely proportional to $1/(2 - g)$. If the convergence criterion were relaxed to, say, 2×10^{-14}, $g = 1.974$ would give convergence after 2300 cycles, a fairly trivial gain, not justified by the loss of precision.

A.3.3.1 Multi-cylinder lenses. To determine the potential distribution in a multi-cylinder lens, we write

$$V_{i,i+1}[z] = V[z + z_{i,i+1}] \qquad \text{and} \qquad E_{i,i+1}[z] = E[z + z_{i,i+1}]$$

where $z_{i,i+1}$ represents the position of the gap between the ith and $(i + 1)$th cylinders. We can then represent the axial potential and its gradient in terms of the electrode potentials as

$$\begin{aligned} V[z] = V_1 &+ (V_2 - V_1)V_{12}[z] & E[z] =&(V_2 - V_1)E_{12}[z] \\ &+ (V_3 - V_2)V_{23}[z] & &+(V_3 - V_2)E_{23}[z] \\ &+ (V_4 - V_3)V_{34}[z] & &+(V_4 - V_3)E_{34}[z] \\ &+ (V_5 - V_4)V_{45}[z] & &+(V_5 - V_4)E_{45}[z] \end{aligned}$$

and then calculate

$$T[z] = E[z]/V[z]$$
$$T2[z] = 0.1875 * T^2[z]$$

where $T2 \equiv T^*$ of chapter 4, and

$$F[z] = V^{1/4}[z]$$

which is needed for conversion from real to reduced radii and *vice versa.*

A.4 Ray tracing for focal data

The screen has a width of 640 pixels and we set this equal to 8 D. We calculate the ray paths at intervals of $D/80$ but display only every fourth point. As the positions of the focal points are not known, we must trace the principal rays starting from points a unit distance from the axis and directed parallel to the axis. The reduced rays are traced in the forward and reverse directions by integrating the Picht equation using Numerov's algorithm (equation (4.3)). This requires two initial values of the reduced radius to be known. These are taken to be $RF[0] = F[0]$ and $RF[1] = F[1]$ for the forward direction, and $RB[639] = F[639]$ and $RB[638] = F[638]$ for the reverse direction. RF corresponds to r_2 of equation (1.5) and figure 1.5 of chapter 1, and RB to r_1. Both pairs of values correspond to a real initial radius of one. Subsequent values of $RF[z]$ are calculated and, if $z\ MOD\ 4 = 0$, a point is plotted to the screen at $(z, RF[z]/F[z])$ with the ordinate suitably scaled to take account of the different resolutions of the various graphics cards. If $RF[z+1]/RF[z] < 0$ then the ray has crossed the axis and the position of the first focal point is found by interpolation. The position of the corresponding principal point is calculated from the slope of the *real* ray, extrapolating back to a radius of 1. The focal length is then found by subtraction. If the ray has not crossed the axis by the end of the calculation (i.e. if $RF[639]$ is still positive), which will happen for weak lenses, the focal properties are all found by extrapolation. Values for the reverse direction are determined in exactly the same manner. The matrix elements are also calculated.

The coefficients of spherical aberration are calculated using equations analogous to equation (5.17) with the integrands expressed by equation (5.18). $RF[z]$ and $RB[z]$ are first scaled to give $R_1[z] = f_1 * RB[z]$ and $R_2[z] = -f_1 * RF[z]$, and the integrals are evaluated using Simpson's rule. Similar calculations, using equations (5.19)–(5.22) give the Petzval integral and the various coefficients of chromatic aberration. The Petzval integrands for the three lenses for which the spherical aberration is illustrated in chapter 5 are shown in figure A.2.

There are certain conditions which will halt the program. If the slope of the ray at the exit plane is zero, and if the ray has not crossed the axis, then there is no lens action. If the ray crosses the axis between the lens centre and the outer electrode, it is treated as being too strong to be useful. If the slope is positive at the exit plane, then the ray must have crossed, or be about to cross, the axis a second time, and the lens is unlikely to be useful.

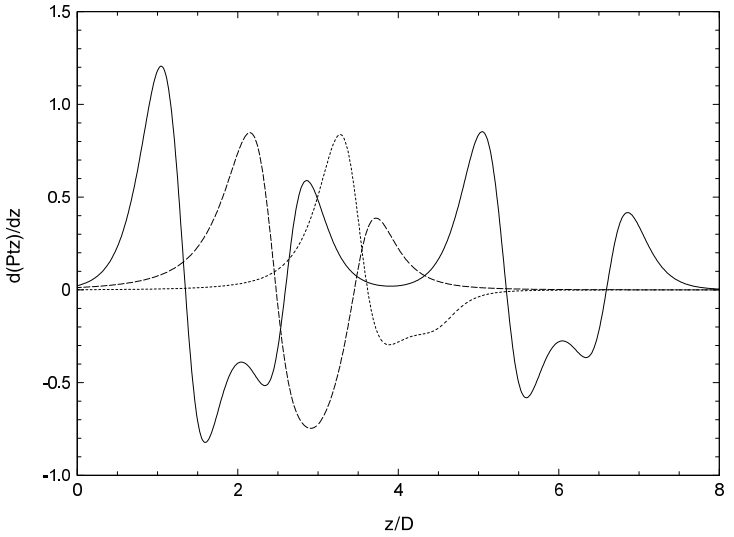

Figure A.2 The Petzval integrands for the three aperture lens (- - - -), the Afocal8 lens (——) and the two aperture lens (– – –) illustrated in figures 5.10 and 5.11.

A.4.1 Accuracy and consistency

The lens data are presented to $3\frac{1}{2}$ significant figures, though the last may not always be too reliable. The calculation of the forward and reverse characteristics are made independently and the ratio of the focal lengths provides a test of the precision of the ray tracing routines. This ratio should equal the square root of the overall voltage ratio. The determinant of the lens matrix should equal $-f_1/f_2$. Further tests may be made by examining lenses which are reversals of each other, for example 10 : 1 and 1 : 10, or 1 : 8 : 2 and 1 : 4 : 0.5 for a three element lens. These tests can also examine the coefficients of spherical aberration, which should be related by equation (5.26).

A.5 Solving the imaging problem

A.5.1 Finding the image

The two initial values of the reduced radius are taken to be $R[0] = 0$ and $R[1] = h * F[1]$, where $h = 1/80$ is the lattice spacing, corresponding to a slope of one for the real radius. The Picht equation is again integrated step by step for the whole length of the system, and each fourth point is plotted, again as real radius. The point where the ray crosses the axis is found in the same way as the focal point and the magnification and the aberration coefficients are calculated as explained later. No rays are traced beyond the limits of the screen and the image point is found by extrapolation if necessary.

Further rays may be traced with initial values $R[0] = r \times F[0]$ and $R[1] = (r + s) \times F[1]$ where r and s are values of the initial radius and slope input in response to the 'Another ray?' prompt.

A.5.2 Solving the lens

The same initial values of the reduced ray are used and traced for the length of the system. The value of $R[I]$, where I represents the image plane, is noted and the focusing voltage, V_f, is adjusted by an amount which depends on this miss distance as

$$V_f = V_f + k \times R[I].$$

After the first trace, k can only be guessed and a value of 0.1 is taken in order that the effect will be quite small. After this first correction, a suitable value for k can be estimated, because $R[I]$ is then known for two values of V_f, and we write

$$k = (V_f - V_f^{\dagger})/(R^{\dagger}[I] - R[I])$$

where the † indicates the *previous* value.

The first change is deliberately made small, because even the sign of k is not initially known. If $R[I]$ is positive, the lens is too weak and must be strengthened. The correct value of V_f may have one of two values: 'high' or 'low', with the 'high' value generally offering better aberration behaviour. However, there are occasions when the 'low' value is of interest. In this case, the appropriate change in V_f is a reduction. The small initial change is a way of letting the program know where the target value of V_f lies.

If the change in V_f leads to a value less than 0.01, it is replaced by a small random value, greater than 0.01. The value is small to keep the search for V_f in the 'low' region, and random so that the program does not immediately get stuck in a loop. The target for satisfactory imaging is that $|R[I]| \leqslant 0.001$.

When the correct value of V_f has been found, the coefficients of spherical and chromatic aberration are calculated using equations (5.17) and (5.20). Two further rays are then drawn from a point above the axis; one is initially parallel to the axis, and the other has a slope of -1 in real space. These two rays cross again at the image plane, and the magnification can be calculated from the final and initial off-axis positions. The exact starting point is chosen to be such that the magnified image lies within the length of the marker drawn perpendicular to the axis at the image plane. This requires that the program knows the magnification before these rays are drawn. The reader should have no difficulty in working out how this is possible!

A.5.3 The afocal lens

This is handled in exactly the same way as any other lens. Though only V_5 is specified by the user, $V_3 = V_5^{1/2}$ and the *two* focusing voltages are related by $V_4 = V_2 \times V_3$. Only the higher values of the focusing voltages are considered.

This is achieved by setting the lower limit for V_2 not at 0.01, but at $\sqrt{V_3}$, a value for which the $V_1-V_2-V_3$ lens is close to minimum power†.

A.5.4 The Varimag lens

Constraints are first applied to the possible values of V_2 ($\geqslant V_3^{1/2}$) and V_4 ($\geqslant (V_3 \times V_5)^{1/2}$). V_2 is set to its minimum value, and V_4 is first set to the greater of V_3 and V_5, and then adjusted to focus the lens. Under these conditions, the magnification of the lens is close to the minimum value possible for the given values of V_3 and V_5, because the $V_1-V_2-V_3$ lens contributes minimally to the total power. To find a value for the magnification close to the maximum possible value, V_4 is set to its minimum value and V_2 set to the greater of V_1, V_3 and adjusted to focus the lens again.

If the target magnification lies outside these limits, the program halts with a message that the maximum (or minimum) value has a certain value. This value is given as some 5% less than the true maximum as a safeguard against later problems if V_3 is changed. If the target magnification is acceptable, then V_4 is set to a value found by interpolating between its values for the minimum and maximum magnification, and V_2 is adjusted to focus the lens. The magnification is compared to the target value and V_4 is re-adjusted. Cycles of focusing using V_2 and magnification adjustment using V_4 continue until the target value is met to within 0.1%. This is quite good enough for any practical purpose, but it is noticeable that the dependence of magnification on the ratio V_4/V_2 is often quite weak.

† The reader might like to examine the focal properties of three element lenses with the centre element at potentials between those of the outer two.

References

[1] Hawkes P W and Kasper E 1989 *Principles of Electron Optics* (London: Academic)
[2] Szilagyi M 1988 *Electron and Ion Optics* (New York: Plenum)
[3] Harting E and Read F H 1976 *Electrostatic Lenses* (Amsterdam: Elsevier)
[4] Picht J 1932 *Ann. Phys., Lpz.* **15** 926–64
[5] Weber E 1939 *J. Appl. Phys.* **10** 663–6
Gray F 1939 *Bell Syst. Tech. J.* **18** 1–18
Bertram S 1940 *Proc. IRE* **28** 418–20
[6] Verster J L 1963 *Philips Res. Rep.* **18** 465–605
[7] Press W H, Flannery B P, Teukolsky S P and Vetterling W T 1986 *Numerical Recipes* (New York: Cambridge University Press)
[8] Read F H 1999 private communication
[9] Renau A, Read F H and Brunt J N 1982 *J. Phys. E: Sci. Instrum.* **15** 347–54
van Hoof H A 1980 *J. Phys. E: Sci. Instrum.* **13** 1081–9
Martinez G and Sancho M 1982 *Am. J. Phys.* **51** 170–4
[10] Edwards D 1983 *Rev. Sci. Instrum.* **54** 1729–35
[11] Natali S, Di Chio D and Kuyatt C E 1972 *J. Res. NBS* A **76** 27–35
[12] Durand E 1957 *C. R. Acad. Sci. Paris* **244** 2355–8
[13] Kasper E 1990 *Nucl. Instrum. Methods* A **298** 295
[14] 1958 *British Association Mathematical Tables* vol 6 (Cambridge: Royal Society)
[15] Bonjour P 1979 *Revue Phys. Appl.* **14** 533–40
[16] Renau A and Heddle D W O 1986 *J. Phys. E: Sci. Instrum.* **19** 284–95
[17] Hawkes P W 1987 *J. Phys. E: Sci. Instrum.* **20** 234–5
[18] Heddle D W O and Kay S M 1996 *J. Phys. B: At. Mol. Opt. Phys.* **29** 1263–71
[19] Heddle D W O 1999 *J. Phys. D: Appl. Phys.* **32** 1447–54

TSMC EX. 1016 Page 139 of 139