THE UNIVERSITY OF LONDON
Imperial College of Science, Technology and Medicine
Blackett Laboratory
Optics Section

The Application of Non Merit Function Based Methods to the Correction of Optical Systems

Christopher Charles Hull

Thesis submitted in partial fulfillment
of the requirements for the degree
Doctor of Philosophy
and for the
Diploma of Imperial College

August 1992

To my parents, brother
and Sally

Abstract

A new non merit function based lens design algorithm has been developed which combines a quadratic approximation to the individual aberration variations with adaptive aberration control techniques. This has resulted in an increase in the magnitude of the applied parameter changes and a reduction in computing load relative to the linear adaptive method. Several examples are given which demonstrate that the new quadratic adaptive correction algorithm can find solutions not readily available to conventional gradient descent methods.

To demonstrate the capabilities of the new algorithm two quite different stages of the design process have been investigated: the generation of starting designs with reduced higher order aberrations and the fine correction of optical systems.

The generation of starting designs with reduced higher order aberrations has entailed the development of a new method for computing higher order aberrations which is explained in detail. The higher order aberrations were reduced using the quadratic adaptive correction algorithm for five initial lens designs which were then optimized using the damped least squares method. Different design forms were found using this technique but these were not consistently superior to the designs resulting from the exclusive use of damped least squares optimization.

The fine correction of optical systems has been attempted by initially determining target values for the third order aberrations which optimize the Strehl ratio and then using the adaptive quadratic correction method to target these aberrations. This technique is described in detail and four examples are used to demonstrate the strengths and limitations of the method.

Acknowledgements

Many people have contributed to this thesis, some quite unknowingly. Although a list of all of them would be impractically long I must single out several to be 'mentioned in dispatches'. First and foremost to my supervisor Jon Maxwell - what I know about lens design is solely due to his efforts over the past three and a half years. All the other members of the optical design group at Imperial College have also greatly contributed: to Pru for many invaluable discussions and help, to Mike for his quiet wit and intelligence, to Charles for forcing me back on to the squash court and to Martha for her interest and questions. I am also very grateful to Robin Smith, Brian Blandford, and Geoff Adams for several discussions and for their gentle criticism of some of the chapters.

Life would not have been the same without my tea time friends: Neil for maintaining his sense of humour even when upside down, Roger who always had the ability to make a dredger look as if it was only scraping the surface, Graham 'mine's a pint' Crofts, Rolf 'one hundred and eighty' Bachen and the new kids on the block Kish and Dave. There are many other past and present members of the optics section who deserve a mention: Roddy Cañas, Andrew West, Andrew Creeke and the original members of 612, the computing intellects of 6M09 (thanks Nic for all the help), all the crew of 631 and Dave Gale (of no fixed abode). To those I have not mentioned by name I also extend my thanks - you have not been forgotten!

I am indebted to the Science and Engineering Research Council for financial support during the course of this work and for the opportunity to travel to the International Lens Design Conference in 1990. I must also thank my current employers, the City University, for having the confidence to employ me without having completed this work and for their tolerance during the final throes of writing. Last but by no means least, it gives me great pleasure to have the opportunity to acknowledge the love, support and encouragement I have received from my family - this thesis is dedicated to you.

Contents

Abstract 1

Acknowledgements 2

Contents 3

List of Figures 7

List of Tables 13

1 Introduction 16

1.1 Historical Overview 16

1.2 Thesis Synopsis 19

1.3 References 23

2 Optimization Methods in Lens Design 27

2.1 Introduction 27

2.2 Algorithms No Longer in Common Usage 28

2.2.1 Relaxation Techniques 28

2.2.2 The Method of Steepest Descent 29

2.2.3 The Optimum Gradient Method 30

2.2.4 The Modified Gradient Method 30

2.2.5 Parabolic Approximation Method 32

2.2.6 The Conjugate Gradient Method 33

2.2.7 Least Squares Minimization 33

2.3 The Damped Least-Squares (DLS) Minimization Algorithm 35

2.4 Enhanced Damped Least Squares Algorithms 42

2.5 Orthonormalisation 44

2.6 The Adaptive Method 47

2.7 Global Optimization 50

2.7.1 Simulated Annealing 50

2.7.2 Other Global Search Techniques 52

2.8 Artificial Intelligence 53

2.9 References 55

3 A Quadratic Adaptive Method for Aberration Correction 62

3.1 Introduction 62

3.2 Mathematical Basis of the Method 64

3.3 Implementation of the Method 68

3.3.1 Solving Equation (3.7) 68

3.3.2 Ill-conditioning 70

3.3.3 Specification of Aberrations during Correction 75

3.3.4 Intractable Aberrations 75

3.3.5 An Illustrative Example of Aberration Control 76

3.3.6 Control of Variables 77

3.3.7 The Quadratic Fitting Step FS 78

3.3.8 Aberration Target Levels 78

3.4 Examples 79

3.4.1 Prmary Aberration Correction of an f/4 Triplet 79

3.4.2 Higher order Astigmatism and Petzval Sum Balancng . . . 82

3.4.3 Further Correction of a 50mm f/4 Triplet 82

3.4.4 Correction of an f/1.7 50mm Double Gauss Lens 83

3.5 Discussion 85

3.5.1 Merit Function Variation During Correction 85

3.5.2 Relative Computational Efficiency 89

3.5.3 Average Step Size 90

3.5.4 Comparison with the Adaptive Method 91

3.6 References 93

4 Generating Starting Designs with Higher Order Aberration Correction 95

4.1 Introduction 95

4.2 Mixed Higher Order Aberrations 96

4.2.1 Mixed Higher Order Monochromatic Aberrations 96

4.2.2 Extension to Chromatic Aberrations 101

4.3 Reduction of the Higher Order Aberrations 101

4.3.1 Simultaneous Correction of the Higher Order Aberrations . 101

4.3.2 The Effect of Boundary Conditions 102

4.3.3 Grouped Correction of Higher Order Aberrations 103

4.3.4 Variation of the Seidel Aberrations 105

4.3.5 Computer Implementation 106

4.3.6 Choice of Variables 106

4.4 Examples 107

4.4.1 Experiments with a 50mm f/4 triplet - design # 1 107

4.4.2 Experiments with a 50mm f/4 triplet - design # 2 112

4.4.3 Experiments with a 50mm f/4 triplet - design # 3 115

4.4.4 Experiments with a 90mm f/2 Petzval lens - design # 4 . . 118

4.4.5 Experiments with a 75mm f/1 IR triplet - design # 5 . . . 121

4.5 References 124

5 The Fine Correction of Optical Systems 125

5.1 Introduction 125

5.2 Wavefront Aberration Approximation to Fifth Order 127

5.3 Testing the Approximate Fifth Order Aberration Coefficients . . 136

5.4 Balancing the Wavefront to Maximize Strehl Ratio 142

5.5 Examples 155

5.6 Discussion 159

5.7 References 160

6 Summary and Conclusions 163

Appendix A – Optimization Nomenclature and Definitions. 170

Appendix B – Proof that $\mathbf{x} = \mathbf{A}^{\mathbf{T}}\mathbf{w}$. 172

Appendix C – Test Lenses from Chapter 3. 173

Appendix D – Test Lenses from Chapter 4. 175

Appendix E – Proof that $Var(W)=Var(W_{even})+Var(W_{odd})$. 180

Appendix F – Test Lenses from Chapter 5. 182

References 186

List of Figures

Figure	Caption	Page
2.1	An illustration of the zig-zag movement during steepest descent optimization.	30
2.2	Variation of the merit function Ψ with respect to the damping factor p during a typical DLS iteration.	37
2.3	Merit function contours for a simple linear system illustrating the effect of changing the damping factor.	38
2.4	Merit function contours for a non-linear system.	39
2.5	Contours of constant merit function for an orthonormal system.	46
2.6	Illustration of 'inner cycles' in Glatzel's adaptive method.	49
3.1	The relationship between the actual aberration variation and the linear and quadratic approximations.	66
3.2	The second order approximation viewed as a smoothing function between points O and P.	68
3.3	A simplified flow chart of the computer program for the quadratic adaptive correction algorithm.	69

Figure	Caption	Page
3.4	Actual aberration variations and approximations for (a) linear variation, (b) predominantly quadratic variation,(c) higher order variation where the QAC algorithm is successful and (d) higher order non-linearities requiring a smaller value of the quadratic fitting step FS.	80
3.5	s & t Curves for an f/4 Triplet (a) before and (b) after flat-field correction.	83
3.6	Unvignetted transverse ray aberrations for the f/4 50mm triplet of example 3.	84
3.7	Unvignetted transverse ray aberrations for the f/1.7 50mm double Gauss derivative of example 4.	85
3.8	(a) Merit function variation during flat-field correction of an f/4 triplet with the QAC algorithm. (b) Merit function variation during flat-field correction of an f/4 triplet with the adaptive method.	87
3.9	(a) Merit function variation during final correction of an f/4 triplet with the QAC algorithm. (b) Merit function variation during final correction of an f/4 triplet with the adaptive method.	88
4.1	Even and odd function contributions to the finite transverse ray aberration.	99

Figure	Caption	Page
4.2	Variation of the mixed higher order aberration M_{III} using u' to control EFL and scaling the lens data.	103
4.3	Variation of the row rates of change of all seven mixed higher order aberrations during correction of an f/2 Petzval lens.	104
4.4	Variation of the Seidel aberrations during combined higher order correction of an f/2 Petzval lens.	105
4.5	Diagrammatic representation of the lenses resulting from the corrections performed on triplet starting design # 1 (see text).	109
4.6	DLS optimization runs on starting design # 1 and higher order corrected designs.	110
4.7	DLS optimization runs on starting design # 2 and higher order corrected designs.	113
4.8	Diagrammatic representation of the lenses resulting from the corrections performed on triplet starting design # 2 (see text).	114
4.9	DLS optimization runs on starting design # 3 and higher order corrected designs.	116
4.10	Diagrammatic representation of the lenses resulting from the corrections performed on triplet starting design # 3 (see text).	117

Figure	Caption	Page
4.11	Diagrammatic representation of the lenses resulting from the corrections performed on the Petzval lens starting design # 4 (see text).	119
4.12	DLS optimization runs on starting design # 4 and higher order corrected designs.	120
4.13	Diagrammatic representation of the lenses resulting from the corrections performed on the IR triplet starting design # 5 (see text).	122
4.14	DLS optimization runs on starting design # 5 and higher order corrected designs.	124
5.1	Finite wavefront aberration values required to approximate the wavefront to a fifth order expansion.	128
5.2	Wavefront aberration component of longitudinal defocus.	130
5.3	Polar diagram showing the variation of elliptical coma with azimuth.	134
5.4	The relationship between distortion and wavefront aberration.	135
5.5	Raytrace computed wavefront aberrations for an f/4 triplet (see text).	137

Figure	Caption	Page
5.6	Wavefront error curves for an f/4 triplet (see text) using finite ray trace results from the margin of the field.	138
5.7	Wavfront error curves for an f/4 triplet (see text) using finite ray trace results from the zone of the field.	140
5.8	Wavfront error curves for a Tessar lens with large edge of aperture aberrations.	141
5.9	Surface contour map of the change in variance, E, as ${}_0w_{20}$ and ${}_0w_{40}$ are varied.	145
5.10	Scatter plot of filtered variance values for different axial aberration balances.	147
5.11	Theoretical wavefront aberrations for an axial aberration balance which maximizes the Strehl ratio.	147
5.12	Theoretical balance of the coma terms in the wavefront aberration to maximize Strehl ratio.	148
5.13	Surface contour profile of the change in variance E as ${}_2w_{20}$ and ${}_2w_{22}$ are varied.	150
5.14	Scatter plot of filtered variance values for different off-axis aberration balances.	151
5.15	Theoretical wavefront aberrations for an off-axis aberration balance which maximizes the Strehl ratio.	152

Figure	Caption	Page
5.16	Typical variation of the Seidel aberrations during Strehl optimization.	154
5.17	Raytrace computed wavefront aberrations for test lens TL5.1 (see text).	155
5.18	Raytrace computed wavefront aberrations for test lens TL5.1 after Strehl optimization.	155
5.19	Spatial frequency response of TL5.1 both before and after Strehl optimization.	156
5.20	Spatial frequency response of TL5.2 both before and after Strehl optimization.	158

List of Tables

Table	Caption	Page
3.1	The condition number, $cond(A)$, of the matrix products $\mathbf{AA^T}$ and $\mathbf{A^TA}$ for six optical change matrices.	71
3.2	A simple correction illustrating the handling of controlled aberrations.	76
3.3	The number of iterations for correction of examples 1–4 using quadratic adaptive correction.	81
3.4	Target values for the flat-field correction of an f/4 triplet (example 2).	82
3.5	Comparison of computational efficiency between the linear adaptive method and the QAC algorithm.	90
3.6	Comparison of the mean step size between the linear adaptive method and the QAC algorithm.	90
3.7	Comparison of the number of iterations between adaptive correction and the QAC algorithm.	92
4.1	Mixed higher order transverse ray aberrations.	100

Table	Caption	Page
4.2	Mixed higher order transverse ray aberrations both before ('start') and after ('pc4-11') combined higher order correction of an f/4 triplet (starting design # 1).	108
4.3	RRC's for starting design # 1.	111
4.4	Mixed higher order transverse ray aberrations both before ('start') and after ('corrected') grouped higher order correction of an f/4 triplet (starting design # 1).	111
4.5	Mixed higher order transverse ray aberrations both before ('start') and after ('corrected') combined higher order aberration correction of an f/4 triplet (starting design # 2).	113
4.6	Mixed higher order transverse ray aberrations both before ('start') and after ('pc4-10') grouped higher order aberration correction of an f/4 triplet (starting design # 3).	116
4.7	Mixed higher order transverse ray aberrations both before ('start') and after ('corrected') grouped higher order aberration correction of an f/2 Petzval lens (starting design # 4).	120
4.8	RRC's during grouped higher order aberration correction of starting design # 5.	121

Table	Caption	Page
4.9	Mixed higher order transverse ray aberrations both before ('start') and after ('pc6-10') grouped higher order aberration correction of an f/2 Petzval lens (starting design # 4).	123
5.1	Third order monochromatic wavefront aberrations.	129
5.2	Approximate fifth order aberration coefficients.	136
5.3	Standard deviations of the approximating wavefronts for the f/4 triplet test lens.	138
5.4	Values for the higher order aberration coefficients in wavelengths for the axial aberration balances shown in figure 5.10.	146
5.5	Values for the higher order aberration coefficients in wavelengths for the off-axis aberration balances shown in figure 5.14.	152

CHAPTER 1

Introduction

1.1 Historical Overview

In 1952 Baker [1.1] was one of the earliest workers to use computers for the iterative improvement of optical systems which was somewhat ambitiously termed "automatic optical design" at that time. Unfortunately his work at Harvard under a US Air Force contract was classified until 1959. The reports subsequently revealed that several of the methods described by other authors in the interim had been evaluated by Baker including an algorithm that was almost identical to the damped least squares algorithm. At about the same time as Baker's work at Harvard, Black was working on univariate methods of lens correction at the University of Manchester [1.2]. This was the first reported work on computer lens optimization in the U.K. and coincided with the publication by R.E. Hopkins of work at the Institute of Optics, University of Rochester, where a computer was used to find the third order solution to a lens system [1.3] by solving an exactly determined set of linear equations. Meiron at the Weizman Institute in Israel and Feder at the Eastman Kodak Company, New York were also working on computer optimization of lens systems and both published a series of papers on the subject in the late 1950's and early 1960's. The methods covered were a parabolic approximation technique [1.4], the optimum gradient method [1.5], the modified gradient method [1.6] and the conjugate gradient method [1.7]. However, all these techniques together with the steepest descent method [1.2], [1.9], suffered from slow convergence and are now considered obsolete.

The least squares method was first used by Rosen and Eldert [1.8] in 1954 as a technique for reducing a single-value merit function which was defined as the sum of the squares of the transverse ray aberrations. The method though has a tendency to oscillate about a solution and sometimes catastrophically diverges from it because of the inherent ill conditioning of optical change matrices. In 1958 Wynne, who was working at Wray Optical Ltd. at that time, described a modification to the least squares algorithm which ed the magnitude of the applied parameter changes and also allowed the solution to continuously vary between a steepest descent increment and a least squares increment [1.9]. The technique was called SLAMS (successive linear approximation using maximum steps) and overcame the limitations of the least squares approach. It was subsequently found to be identical to the work of Levenberg who had developed the method to find least squares solutions to certain non-linear problems [1.10]. Wynne's clarity of explanation made the results from his work readily accessible and SLAMS rapidly grew in popularity. Under its new name of the damped least squares algorithm (DLS) it has become the most widely used and researched optimization algorithm for lens design and is available in almost every computer-aided optical design program.

In the early 1960's with computers becoming more powerful and increasingly widespread in their availability, the amount of work on computer-aided lens design increased. Glatzel, who was working for Carl Zeiss at Oberkochen, first published his adaptive method in 1962 in the German journal Optik [1.11] although an English translation was not available until the Conference on Lens Design with Large Computers which was held at the University of Rochester in 1966 [1.12]. Glatzel's work received little attention although R.E. Hopkins and Spencer at the University of Rochester had been working along similar lines and published their results in the same year [1.13]. Grey's orthonormalization optimization method was published in 1963 [1.14] and it presented an approach that attempted to counter some of the major problems in lens optimization namely non-orthogonality of the aberrations with respect to the parameter base and ill-conditioning of the matrices. Grey's method together with damped least squares and Glatzel's adaptive method are the only optimization algorithms from the early work on lens optimization which have

survived to the present day.

Most of the work during the late 1960's and 1970's centred around the damped least squares algorithm and considered such issues as selection of the damping factor [1.15], the use of different merit functions [1.16] and convergence [1.17]. It was not until desk top computers became available in the late 1970's that the direction of computer lens design software and optimization methods significantly changed. Initially, desktop computers (later to become personal computers, or PC's) had only limited memory and great emphasis was placed on computational efficiency and program size. This further encouraged the development of efficient numerical methods for lens design, particularly during optimization where expensive function evaluations could make an iterative method impractically slow. However, arguably the most significant outcome of the desk top computer revolution has been the vast increase in the number of people who have access to lens design software. Optimization programs now need to handle multi-configuration zoom systems, non-rotationally symmetric systems, many-colour optimization, direct optimization of the spatial frequency response, optimization of novel optical elements such as diffraction gratings, GRIN lenses .... the list is endless. Many of the techniques though were developed for large mainframe computer programs and the challenge was to make computer programs small and efficient enough to fit on a PC. In addition, the user interface was of increasing importance, particularly for those optical engineers who did not spend every day using a program.

The development of personal computers was so rapid that the problems of computer speed and program size were shortlived and in particular computer processing power increased to such an extent that by the mid 1980's so called global optimization became feasible. Some of the earliest attempts at global optimization used the generalized simulated annealing algorithm and were reported in 1987 by Hearn [1.18] and Weller [1.19]. More recently Forbes and Jones [1.20] have continued working in this area although the initial enthusiasm of many lens designers appears to have subsided. The problem with any global optimization method is the 'curse

of dimensionality'; for example, if all six curvatures of a triplet are allowed to vary and there are 10 possible values for each curvature then 10^6 system evaluations are required to assess all the possible designs. One of the most recent examples from Forbes and Jones [1.21] was a four element lens which required their adaptive simulated annealing algorithm to run for 8 hours on a UNIX workstation. More recently Sturlessi and O'Shea [1.22] have described a parametric survey technique which covers the entire parameter space with a uniformly spaced grid and evaluates the system at each point on the grid. Again this is computationally very expensive and more recent work by these authors described using the technique as a method for finding likely regions of parameter space for local minimization algorithms. One of the main reasons that these algorithms should not be dismissed as impractical is that a majority of PC's and workstations are idle for at least half of the day. Therefore to set a global optimization program running overnight does not involve any loss of resources and may at best find a better solution or at the worst be able to suggest some alternative configurations.

In parallel with the work on global optimization was the application of artificial intelligence and expert systems to lens design. This was first discussed by Bohachevsky *et al.* in 1984 [1.23] and has received a steady increase in interest since then. There has however, only been one instance, to the author's knowledge, where an artificially intelligent program has been implemented in commercial software and where examples of its application have been given [1.24]. However, ambitious projects are currently under development, most notably at the University of Alabama [1.25], which may yet prove successful.

With the widespread use of computer graphics it has become common to study the state of correction of a lens system through the graphical display of aberrational information rather than by the direct use numerical ray trace results. The effects of the classical aberrations on the transverse ray error curves and the optical path difference curves are therefore well known to all practising lens designers. Aberrations that need to be corrected are usually identified through the finite transverse

ray errors resulting from ray trace calculations. However, each transverse ray error contains information about many different aberrations and to all orders of approximation. But such techniques have achieved a large measure of success.

Much of the work over the past thirty years in the field of lens optimization has involved variations of the least squares method where there are always more aberrations to correct than variable parameters. In these methods it is necessary to define a single figure of merit which can be reduced as in a least squares linefit where the sum of the squares of the deviations is minimized. The merit function that has been most extensively used in lens design is the sum of the squares of the transverse ray errors although the merit function may use wavefront aberrations and in addition, will usually contain other quantities including paraxial constraints and boundary violations.

One of the problems with the least squares approach is that valuable information is lost about the distribution of the magnitudes between different aberrations. This is unfortunate since most lens designs are assessed in terms of the aberrations that are limiting the performance of the system and which need to be altered to achieve a 'better' aberration balance. In the least squares environment this is achieved indirectly by altering the contributions to the merit function and re-optimizing the system. This procedure may have to be repeated several times to achieve the desired effect assuming it is possible.

The merit function variation defines a surface in n-dimensional parameter space which can be considered similar in terrain to a mountainous landscape with long winding valleys and many local depressions. Standard gradient algorithms only have the ability to move in a direction in which the merit function decreases. They can therefore become trapped by local minima and are also dependent on the starting point for the descent (the initial design). This limitation has been recognized and many variants of the basic DLS algorithm are now available which allow the user to define his/her own merit function and to target aberrations to specified values. The

use of different merit functions has been prompted to help avoid local minimum solutions by manipulating, in some ill-defined sense, the topography of the merit function surface. It is also becoming increasingly common for programs to use a merit function based on a direct measure of image quality. This may be, for example, the variance of the wavefront aberration or the optical transfer function.

1.2 Thesis Synopsis

This thesis is concerned with methods for correcting optical systems which do not use a single-value merit function. If a merit function is not defined and if instead individual aberration values are targeted, it is possible to find solutions that are not readily available to gradient methods that use a merit function. Following from the discussions in the previous section, such a method can provide the designer with direct control of a design since individual aberrations can be adjusted rather than having to go through the less direct route of changing optimization parameters and re-optimizing the merit function. However, the drawback with such methods is that the user must specify target values for the aberrations. Although this is possible with experience, it is tedious and may require more than one optimization run to reach the desired result; it should be possible to avoid such an indirect approach.

Historically, the ideas for lens correction in this thesis evolved from a project that intended to use 'optical intelligence', or information regularly used by lens designers to create an expert systems type of environment for lens correction. Such an environment would use standard gradient optimization methods where appropriate but not during final aberration balancing or during the initial correction of a lens that is far away from a suitable solution. In addition different correction strategies would be programmed to attempt the correction of aberrations that have been diagnosed as fundamentally limiting the design from further improvement. Such techniques required an algorithm that was much more flexible than a standard DLS

optimization program with the ability to handle composite variables, any type of aberration and to be able to specify suitable target values for the aberrations.

Chapter 3 of this thesis describes a lens correction algorithm which is a hybrid of Glatzel's adaptive method [1.26] and a quadratic extrapolation method described by Maxwell and Hull [1.27]. It combines many of the aberration control techniques used in the adaptive method together with the ability to take large steps in parameter space due to the second order approximation to the aberration variation. The quadratic approximation also allows the explicit indication of intractable aberrations and the precise targeting of aberrations within the region of a solution. The algorithm is general and can be applied to the correction of optical systems where transverse ray aberrations or aberration coefficients are targeted.

Chapter 4 applies this algorithm to the reduction of higher order aberrations during the initial stages of the design process. Higher order aberrations vary more slowly than third order aberrations and hence, if they can be reduced at an early stage of the design process, the final aberration balancing has a greater potential for success. Conventional correction methods are ill-suited to reducing the higher order aberrations since the linear approximation to the aberration variation restricts the size of the applied parameter changes and more importantly, when the starting design is a long way from a good solution, it is likely that the design will converge to an inferior local solution. The reduction of the higher order aberrations considered in chapter 4 is an attempt to find new starting designs in a manner which is independent of a merit function and which may then converge to better solutions than is possible using conventional methods.

Chapter 5 presents an approach for optimizing the Strehl ratio of an optical system by adjusting only the third order aberrations during the final stage of aberration correction using the algorithm described in chapter 3. The higher order aberrations remain approximately static for the small changes required to adjust the third order aberrations and therefore the algorithm converges rapidly. The higher order aber-

rations are found by approximating the wavefront up to and including the seventh order spherical aberration term and subtracting the Seidel aberrations from finite ray trace computed wavefront aberrations. Similar approaches have been given by Cox [1.28] and King [1.29]. The derivation of the third order coefficients which balance their higher order counterparts to optimize the Strehl ratio is given in some detail. Several observations are also made about the variance of the wavefront aberration. The technique is suitable for systems with relatively small aberrations where wavefront aberrations are an appropriate method of assessing image quality.

Chapter 6 summarises the main results and draws conclusions about the advantages and limitations of applying the methods described in the earlier chapters.

1.3 References

[1.1] J.G. Baker, Technical Reports 1–12 and a Supplement prepared by Perkin-Elmer Corp., under Air Force Contract AF 33(038)–10836. Declassified 1959.

[1.2] G. Black, "On the Automatic Design of Optical Systems," Proc. Phys. Soc, B, **68,** 729–736 (1955).

[1.3] R.E. Hopkins *et al.*, "Automatic Correction of Third-Order Aberrations ", J.O.S.A., **45**, 363–365 (1955).

[1.4] J. Meiron and G. Volinez, "Parabolic Approximation for Automatic Lens Design," JOSA, **50**, 207–211 (1960).

[1.5] D.P. Feder, "Automatic Lens Design Methods," J.O.S.A., **47**, 902–912 (1957).

[1.6] J. Meiron and H.M. Loebenstein, "Automatic Correction of Residual Aberrations," JOSA, **47**, 1104–1109 (1957).

[1.7] D.P. Feder, "Automatic Lens Design with a High-speed Computer," JOSA, **52**, 172–176 (1962).

[1.8] S. Rosen and C. Eldert, "Least-squares Method for Optical Correction," J.O.S.A., **44**, 250–252 (1954).

[1.9] C.G. Wynne, "Lens Designing by Electronic Digital Computer I," Proc. Phys. Soc., **73**, 777–787 (1959).

[1.10] K. Levenberg, "A Method for the Solution of Certain Non-linear Problems in Least-squares," Quart. J. Appl. Math., **2**, 164 (1944).

[1.11] E. Glatzel, "Ein neues Verfahren zur automatischen Korrektion optischer Systeme mit elektronischen Rechenmaschinen," Optik, **18**, 577–580 (1961).

[1.12] E. Glatzel, "A New Method for the Automatic Correction of Optical Systems Using Electronic Computers", in Proc. Conf. *Lens Design with Large Computers*, Rochester, NY, USA, 5–8 July 1966, 23-41—23-47.

[1.13] R.E. Hopkins and G.H. Spencer, "Creative Thinking and Computing Machines in Optical Design," J.O.S.A., **52**, 172–176 (1962).

[1.14] D.S. Grey, "Aberration Theories for Semi-automatic Lens Design by Electronic Computers I. Preliminary Remarks," JOSA, **53**, 672–676 (1963).

[1.15] D.R. Buchele, "Damping Factor for the Least-squares Method of Optical Design," Appl. Opt., **7**, 2433–2435 (1968).

[1.16] W.B. King, "Use of the Modulation-Transfer Function (MTF) as an Aberration-Balancing Merit Function in Automatic Lens Design," J.O.S.A., **59**, 1155–1158 (1969).

[1.17] D.C. Dilworth, "Pseudo-second Derivative Matrix and its Application to Lens Design," Appl. Opt., **17**, 3372–3375 (1978).

[1.18] G.K. Hearn, " Generalized Simulated Annealing Optimization Used in Conjunction with Damped Least-squares Techniques" in Proc. SPIE Vol. **766**, *Recent Trends in Optical System Design*, 283–284 (1987).

[1.19] S.W. Weller, "Simulated Annealing: What good is it?" in Proc. SPIE Vol. **818**, *Current Developments in Optical Engineering II*, 265–274 (1987).

[1.20] G.W. Forbes and A.E.W. Jones, "Towards Global Optimization with Adaptive Simulated Annealing", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, 144–153 (1990).

[1.21] G.W. Forbes and A.E.W. Jones, "Global Optimization in Lens Design", Optics and Photonics News, 23–29, March 1992.

[1.22] D. Sturlessi, and D.C. O'Shea, "Future of Global Optimization in Optical Design", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, 54–68 (1990).

[1.23] I. Bohachevsky, V.K. Viswanathan, and G. Woodfin, "An 'Intelligent' Optical Design Program," in Proc. SPIE Vol. **485**, *Applications of Artificial Intelligence*, 104–112 (1984).

[1.24] D.C. Dilworth, "Applications of Artificial Intelligence to Computer-Aided Lens Design," in Proc. SPIE Vol. **766**, *Recent Trends in Optical System Design: Computer Lens Design Workshop,* 91–99 (1987).

[1.25] A. Choudry, and R.B. Johnson, "Knowledge Based Environment for Optical System Design," in Proc. SPIE Vol. 805, *Optical Components and Systems,* (1987).

[1.26] E. Glatzel and R. Wilson, "Adaptive Automatic Correction in Optical Design," Appl. Opt., **7**, 265–276 (1968).

[1.27] J. Maxwell and C.C. Hull, "A Multidimensional Quadratic Extrapolation Method for the Specific Correction of Aberrations in Optical Systems," Appl. Opt., **31**, 2234–2240 (1992).

[1.28] A. Cox, *A System of Optical Design*, Focal Press, London (1964).

[1.29] W.B. King, PhD Thesis, University of London (1966).

CHAPTER 2

Optimization Methods in Lens Design

2.1 Introduction

This chapter presents a review of many of the optimization techniques that have been applied to lens design. Its purpose is to clarify the current state of work in this field and to put into context the material described in chapters 3–5 of this thesis. It includes algorithms that no longer appear to be in current use and the more recent investigations into global optimization and artificial intelligence.

There is a large amount of literature on optimization theory and the methods for 'automatic' optical design. Few authors, however, have attempted to compare the various algorithms and Takahashi [2.1] is the only one who, to the author's knowledge, has presented a practical comparison based on the optimization of model functions. Two possible reasons for this are that consistency in programming methods is required and that there are numerical difficulties associated with the practical implementation of some algorithms. It is known, for example, that Grey's orthonormalization method will fail if it is implemented without the numerical safeguards which have been discussed by Grey [2.2]. Consequently, such a study would be very time consuming and its value questionable in some respects.

2.2 Algorithms No Longer in Common Use

2.2.1 Relaxation Techniques

One of the earliest reported attempts at computer-aided lens design and optimization was carried out by Black [2.3] at the University of Manchester in the early 1950's. The method he described successively alters each available design parameter in an attempt to minimize a system merit function, Ψ, which is defined as the sum of the squares of the residual aberrations. When a change in one variable does not improve the merit function, the minimum for that particular parameter has been reached and control moves on to the next parameter. Such a technique produces a rapid improvement over the first few changes but the process rapidly stagnates. Three methods were described to 'upset' a system in this condition. Firstly, block operations can be used where two or more variables are changed simultaneously; bending is an example of a block operation. Secondly, it is possible to use a group operation to reduce Ψ where extrapolation is applied to two or more systems to produce a system with an even lower merit function. This requires that a large number of system specifications, together with their calculated residuals, are held in memory at any one time. Finally, when none of these processes affects any further improvement in the merit function, the system can be subjected to a random operation. This is a random application of changes of random length applied to the system parameters. There is a connection here with Monte Carlo methods which are now used for lens optimization on some of the larger computers [2.4].

The four techniques described above were run in parallel, the latter three being used to upset the design before returning it to the variable by variable minimization. The convergence rate was extremely slow, (Black presented a Tessar lens that had undergone two hours of minimization on a computer that could calculate 16 aberration residuals in a minute), and the method lacks consistency. However, these techniques present an interesting historical perspective when more recent work on global optimization is considered (see §2.7).

2.2.2 The Method of Steepest Descent

The steepest descent algorithm was first proposed by Cauchy [2.5] as a method for solving sets of simultaneous linear equations. Several authors ([2.3, 2.6–2.8]) have since described its application to lens design as a method for lens optimization. It is also known as the gradient method.

The steepest descent method proceeds as follows. A merit function is defined as the sum of the squares of the aberrations and the vector of first order partial derivatives of the merit function with respect to the system parameters, $\nabla\Psi$, is calculated. The direction of $\nabla\Psi$ is normal to a surface on which Ψ is constant and therefore movement in the $-\nabla\Psi$ direction will reduce the value of the merit function. The incremental changes in the design parameters are therefore chosen to be proportional to $\frac{\partial\Psi}{\partial x_j}$ and this vector of parameter changes multiplied by a scalar step-length h. Selection of the step-length is important since ideally it should be chosen to reduce the merit function to zero (if the higher derivatives are also zero). Too large a step length can result in an increase in Ψ due to its quadratic form and non-linearities, whereas too small a value of h results in very slow convergence. The calculation of the optimum value of h will be considered in the next section.

The rate of convergence of the steepest descent algorithm is too slow to be of any practical value in practical optimization problems. The reasons for this have been most clearly explained by Wynne [2.7]. For a corrected system, the aberration residuals are usually small, resulting in small steps towards the solution. In addition, the algorithm requires that simultaneous changes to *all* the parameters produce a decrease in the merit function. However, in a corrected system, the most likely outcome of a parameter change is an *increase* in Ψ. Consequently the parameter changes are very small and progress towards the minimum is very slow. If the aberrations, f_i, are linear with respect to the system parameters, x_j, then the contours of constant Ψ define a series of concentric hyper ellipsoids. The algorithm produces a zig-zag movement which is rarely directed towards the centre because of

the extreme eccentricity of the ellipses (figure 2.1). In addition, any non-linearity in the aberrations tends to warp and twist the ellipsoids making the situation even worse (see §2.3).

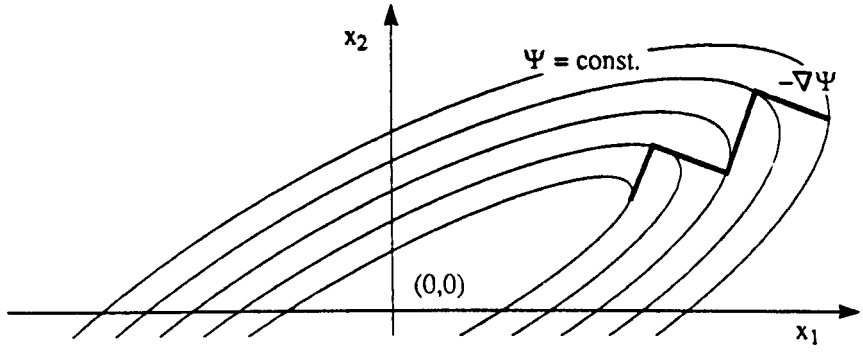


Figure 2.1 An illustration of the zig-zag movement during steepest descent optimization.

Mathematically, the slow convergence of the steepest descent method can be theoretically proved to be caused by the difference in the sensitivities of the parameters and not the non-linearities of the aberrations [2.9].

2.2.3 The Optimum Gradient Method

The optimum gradient method is an attempt to accelerate the convergence of the steepest descent algorithm and was first described by Feder [2.6]. At each iteration the step length h is calculated such that $\Psi(h)$ is a minimum. This can be achieved in one of two ways: Ψ can be evaluated for several values of h and a polynomial fitted to the data. It is then a simple matter to find h_{min}. Alternatively $\nabla\Psi$ can be evaluated for several values of the step-length and the scalar product $\nabla\Psi(h)\cdot\nabla\Psi(0)$ plotted against h. This function is zero for stationary values of $\nabla\Psi(h)$ (which of course does not rule out the possibility of finding points of inflection) and occurs when $\nabla\Psi(h)$ is perpendicular to $\nabla\Psi(0)$ explaining the typical zig-zag movement

towards the solution. In the optimum gradient method the ratio $\frac{h}{h_{min}}$ is always chosen to be 1. If this ratio is less than unity then one is said to undershoot the minimum and if $\frac{h}{h_{min}} > 1$, then the minimum is overshot.

Although the optimum gradient method is simple and general, it is also very slow even when applied to linear equations which can be readily solved by Gaussian elimination.

2.2.4 The Modified Gradient Method

Meiron and Loebenstein [2.8] first described this method which is an adaptation of of the method of steepest descent to try and improve convergence. For small increments in the design parameters, Δx_j, the change in the merit function is given to a first order approximation, by:-

$$\Delta\Psi = \sum_{j=1}^{n} \left(\frac{\partial\Psi}{\partial x_j}\right) \Delta x_j \tag{2.1}$$

and the changes Δx_j are constrained to a small region by the condition:-

$$\sum_{j=1}^{n} (\rho_j \Delta x_j)^2 \leq \varepsilon \tag{2.2}$$

where the weighting factors, ρ_j, are introduced approximately to equalize the different variabilities of the parameter changes. Although the values of ρ_j vary over large parameter changes, they can be regarded as constant within a small region. Inequality (2.2), therefore, restricts the 'distance' moved in parameter space.

Equation (2.1) is linear in Δx_j and hence the largest decrease in the merit function occurs on the boundary of the valid region specified by (2.2). As a result (2.2) can be written as an equality. The values of Δx_j are then calculated using Lagrange multipliers to give

$$\Delta x_j = -\frac{k}{\rho_j^2} \frac{\partial\Psi}{\partial x_j} \tag{2.3}$$

where k is a positive coefficient of proportionality. Increasing k in steps, decreases Ψ along a path which can be considered to be of steepest descent. The formulation of

this method using Lagrangian multipliers also allows the handling of additional constraints; for example, maintaining the back focal length or restricting the steepness of the curvatures.

Although two examples are presented which demonstrate the mathematical validity of the algorithm, no mention is made of efficiency or convergence speed. Takahashi [2.1] noted that the convergence was better than any of the techniques discussed in the previous sections but was still rated as 'poor.'

2.2.5 Parabolic Approximation Method

Meiron and Volinez [2.11] have described another modification to the gradient method where the variation of the merit function, Ψ, with each parameter, x_j, is approximated by a parabola. Three values of Ψ are calculated to determine the parabola: $x_j - \Delta x_j$, x_j and $x_j + \Delta x_j$ where Δx_j is small. The minimum of the parabola is found computationally by replacing the derivatives in the derived mathematical equation for $(x_j)_{min}$ by finite difference approximated derivatives. The value of the parameter change to give the minimum value of Ψ is determined for each parameter in turn while holding the other parameters constant at their initial values. The computed values of all the Δx_j's are then added to the variables and the new system evaluated.

The algorithm suffers from the same problem as the optimum gradient algorithm of Feder [2.6], namely that more computation is required at each stage when compared with the techniques discussed earlier in this section. Therefore, although the rate of convergence is faster, this is to some extent nullified by the increased computation time. However, a 'better' minimum is often found because the separate minimization of each parameter in the parabolic approximation method is not so adversely affected by an ill-conditioned optical system. Consequently, satisfactory solutions are more likely to be found.

2.2.6 The Conjugate Gradient Method

The conjugate gradient method was originally described by Hestenes and Stiefel [2.12] but first appears in the optical literature some ten years later [2.13], [2.10]. Although the mathematics are elegant, it is doubtful whether any advantage is gained over Gaussian elimination for solving sets of linear equations. However, with 'slight modifications,' Feder [2.6] applied it successfully to optical design.

In the steepest descent method, a step was made in the $-\nabla\Psi$ direction to minimize Ψ. In the conjugate gradient method, a correction vector, which is dependent upon $-\nabla\Psi$ *at the previous iteration*, is added to the gradient vector. It has been mathematically proved [2.12] that for N linear equations, the technique converges in N steps or less – a significant improvement over the steepest descent method where the number of steps may be raised to the power 6 or 7! In optical design, Feder [2.10] states that 'the important thing is that this gain in speed carries over to the non-linear case.' However, Takahashi [2.1] finds no significant improvement in speed of convergence and claims a faster convergence for the parabolic approximation method of Meiron and Volinez [2.11].

2.2.7 Least Squares Minimization

In lens design it is common to have more aberrations to control than available variable parameters. Such situations lend themselves to a least squares approach which finds the "best compromise" solution to the problem. It is important to consider the conditions for obtaining a least squares solution and the necessity for defining a merit function (or, perhaps more correctly an error function). Therefore, the least squares normal equation will be derived in some detail. Discussions and definitions of some of the optimization nomenclature are given in Appendix A.

The aberrations of an optical system, f_i, are continuous functions of the system

parameters therefore they can be expanded in a Taylor series about the current aberration value

$$f_i = f_o + \sum_{j=1}^{N} \frac{\partial f_i}{\partial x_j} x_j + \ldots \tag{2.4}$$

where f_o is the aberration value for system parameters $\mathbf{x_o}$ and x_j are the parameter changes. The merit function Ψ, is formed from the sum of the squares of the aberrations f_i such that

$$\Psi(\mathbf{x}) = \sum_{i=1}^{m} f_i^2. \tag{2.5}$$

Apart from the numerical convenience of having a sum of squares merit function where negative aberrations do not cancel out their positive counterparts, there are sound mathematical reasons for the definition of Ψ given in equation (2.5) which will become apparent.

Equation (2.4) can be written in matrix/vector notation as

$$\mathbf{f} = \mathbf{f_o} + \mathbf{A}.\mathbf{x} \tag{2.6}$$

where $\mathbf{A}$ is the $m \times n$ matrix of partial differentials with elements $A_{ij} = \frac{\partial f_i}{\partial x_j}$. The merit function to be minimized becomes†

$$\begin{aligned} \Psi(\mathbf{x}) &= \|\mathbf{f_o} + \mathbf{Ax}\|^2 \\ &= (\mathbf{f_o} + \mathbf{Ax}) \cdot (\mathbf{f_o} + \mathbf{Ax}) \\ &= \mathbf{f_o} \cdot \mathbf{f_o} + 2\mathbf{f_o} \cdot \mathbf{Ax} + \mathbf{Ax} \cdot \mathbf{Ax} \\ &= \mathbf{f_o} \cdot \mathbf{f_o} + 2\mathbf{A^T f_o} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{A^T Ax} \end{aligned} \tag{2.7}$$

and hence the gradient vector, $\nabla\Psi(\mathbf{x})$, is found by differentiating (2.7) with respect to $\mathbf{x}$

$$\nabla\Psi(\mathbf{x}) = 2\mathbf{A^T f_o} + 2\mathbf{A^T Ax}. \tag{2.8}$$

Differentiating again, the Hessian $H\Psi(\mathbf{x})$ is obtained;

$$H\Psi(\mathbf{x}) = 2\mathbf{A^T A}. \tag{2.9}$$

† The following relations have been used in deriving (2.7): $\mathbf{u}\cdot\mathbf{v}=\mathbf{u^T v}$ and $(\mathbf{AB})^T=\mathbf{B^T A^T}$

In general the matrix product $\mathbf{A}^T\mathbf{A}$ is positive definite and hence the Hessian $H\Psi(\mathbf{x})$ is also positive definite. As a result $\Psi(\mathbf{x})$ is a convex function and therefore has a unique minimizer at the point $\mathbf{x}^*$ for which $\nabla\Psi(\mathbf{x}^*) = 0$ (see Appendix A). The normal equation which characterizes $\mathbf{x}^*$ is given by

$$\mathbf{A}^T\mathbf{f}_o = -\mathbf{A}^T\mathbf{A}\mathbf{x}^*. \tag{2.10}$$

Since the matrix $\mathbf{A}^T\mathbf{A}$ is positive definite it can be inverted, to give

$$\mathbf{x}^* = -\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{f}_o. \tag{2.11}$$

The two mathematical reasons for defining the merit function as the sum of the squares of the aberrations are now apparent; they are that $\Psi(\mathbf{x})$ is a convex function of $\mathbf{x}$ and that $\Psi(\mathbf{x})$ is also twice differentiable with respect to $\mathbf{x}$. Both of these features are essential in deriving the normal equation (2.10).

Least-squares was one of the first minimization techniques to be applied to lens design [2.14], [2.6], [2.15]. However, it proved unsuccessful for two main reasons; the matrix $\mathbf{A}$ and hence the matrix product $\mathbf{A}^T\mathbf{A}$ is prone to ill-conditioning and the method has a tendency to oscillate about and sometimes diverge catastrophically from a solution [2.10]. The damped-least squares algorithm described in the next section overcomes these difficulties.

2.3 The Damped Least-Squares (DLS) Minimization Algorithm

Levenberg [2.16] first described a method for applying least squares to solving certain non-linear problems. The same method was discovered independently by Wynne [2.7] who presented the first results on its application to lens design. The two equations to be normalised to form the damped least squares solution are

$$\mathbf{f}_o + \mathbf{A}\mathbf{x} = 0 \tag{2.13}$$

$$p\mathbf{x} = 0 \tag{2.14}$$

where p is known as the damping factor because it limits the magnitude of the solution vector. Following an identical approach to that in the previous section we have that

$$\begin{aligned}\Psi(\mathbf{x}) &= \|\mathbf{f_o} + \mathbf{A}\mathbf{x}\|^2 + \|p\mathbf{x}\|^2 \\ &= (\mathbf{f_o} + \mathbf{A}\mathbf{x}) \cdot (\mathbf{f_o} + \mathbf{A}\mathbf{x}) + p\mathbf{x} \cdot p\mathbf{x} \\ &= \mathbf{f_o} \cdot \mathbf{f_o} + 2\mathbf{f_o} \cdot \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{x} \cdot \mathbf{A}\mathbf{x} + p\mathbf{x} \cdot p\mathbf{x} \\ &= \mathbf{f_o} \cdot \mathbf{f_o} + 2\mathbf{A}^{\mathrm{T}}\mathbf{f_o} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} + p^2\mathbf{x} \cdot \mathbf{x} \end{aligned} \tag{2.15}$$

and differentiating equation (2.15) with respect to $\mathbf{x}$ gives the gradient of Ψ

$$\nabla\Psi(\mathbf{x}) = 2\mathbf{A}^{\mathrm{T}}\mathbf{f_o} + 2\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} + 2p^2\mathbf{x}. \tag{2.16}$$

The normal equation that characterizes the damped least squares algorithm is found by equating the gradient vector to zero

$$(p^2\mathbf{I} + \mathbf{A}^{\mathrm{T}}\mathbf{A})\mathbf{x}^* = -\mathbf{A}^{\mathrm{T}}\mathbf{f_o} \tag{2.17}$$

and hence the required change in the system parameters is given by

$$\mathbf{x}^* = -(p^2\mathbf{I} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{f_o} \tag{2.18}$$

where $\mathbf{x}^*$ defines a DLS solution step. In optical design the damping factor serves two purposes; it restricts the magnitude of the applied parameter changes to the region over which a linear approximation is valid (equation (2.4)) and it also prevents the matrices becoming singular as a result of ill-conditioning. This is because the damping factor adds a positive value to all terms on the diagonal of the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and hence even if one diagonal element becomes zero due to $\mathbf{A}$ being singular, a solution is still possible. The dual purpose of p has been criticised by some authors [2.17] who have developed methods of eliminating the ill-conditioning based on orthogonalising the aberrations. The subject of orthogonalization will be considered in more detail in the §2.5 where Grey's orthonormalization optimization method is discussed.

A variation of the merit function Ψ with respect to p during a typical iteration is shown in figure 2.2. At each iteration there is an optimum value of p which has

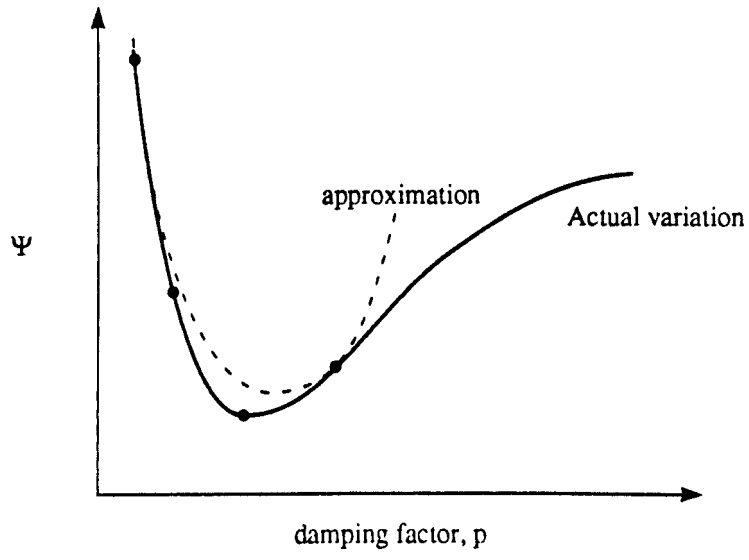


Figure 2.2 Variation of the merit function Ψ with respect to the damping factor p during a typical DLS iteration.

to be found. Levenberg [2.16] proposed a linear search which was time consuming and has been superceded by a method due to Marquardt [2.18] where from an initial arbitrary value, the damping factor is multiplied or divided by a factor υ depending on whether the merit function is higher or lower than its predecessor. (The suggested value for υ was 10 but at least one optical design program uses a value of 2 [2.19]). Fletcher [2.20] has indicated inefficiencies in the Marquardt method for a poor starting value of the damping factor and has proposed a method where the starting value for the damping factor is governed by R, the ratio between the linearly predicted decrease and the actual value of the merit function. If $R \approx 1$ then the first-order approximation to the Hessian, $H(\Psi) = 2(\mathbf{A}^{\mathrm{T}}\mathbf{A})$, used in the least squares increment (equation (2.11)) is good and p can be decreased for the next iteration. However, if $|1 - R| > 0.3$, then the approximation is poor and p must be increased.

The DLS algorithm is known as a variable metric method [2.21] because the damping factor allows a smooth transition between a steepest descent increment and the least squares (or Gauss-Newton) solution:-

$$\lim_{p \to \infty} \mathbf{x}^* = -\mathbf{A}^{\mathrm{T}}\mathbf{f}_o/p^2 \tag{2.19}$$

$$\lim_{p \to 0} \mathbf{x}^* = -\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{f}_o. \tag{2.20}$$

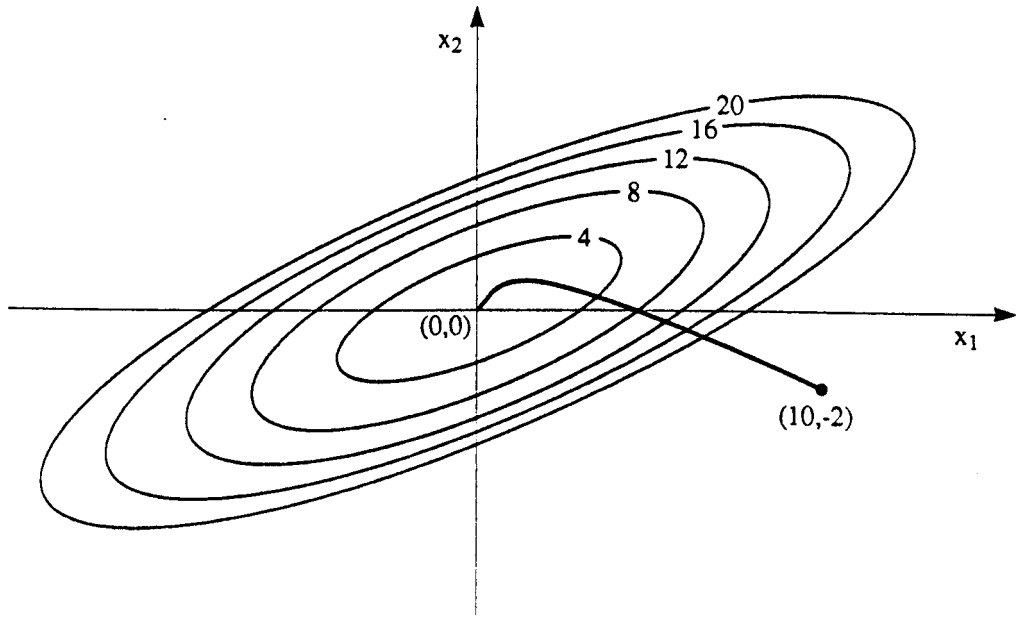


Figure 2.3 Merit function contours for a simple linear system illustrating the effect of changing the damping factor.

The effect of the damping factor can be viewed for the simple two dimensional case if two arbitrary 'aberrations' are defined which are functions of the two variable parameters x_1 and x_2. These aberrations do not relate to any real optical system and have been constructed simply for the purpose of examining the damped least squares solution. Figure 2.3 shows the merit function contours at values of $\Psi = 4, 8, 12, 16, 20$ for the simple linear system

$$f_1 = -0.8x_1 + 0.2x_2 \tag{2.21}$$

$$f_2 = -0.4x_1 + 0.4x_2 \tag{2.22}$$

consisting of two aberrations, f_1 and f_2, which are linear functions of the two variables x_1 and x_2. To illustrate the effect of changing the damping factor, a starting point of $(10, -2)$ has been taken and the DLS equation (2.18) solved for values of $p = 0$ to 10. The values of x_1 and x_2 have been plotted after applying the change and are represented on figure 2.3 by the curved line joining the minimum

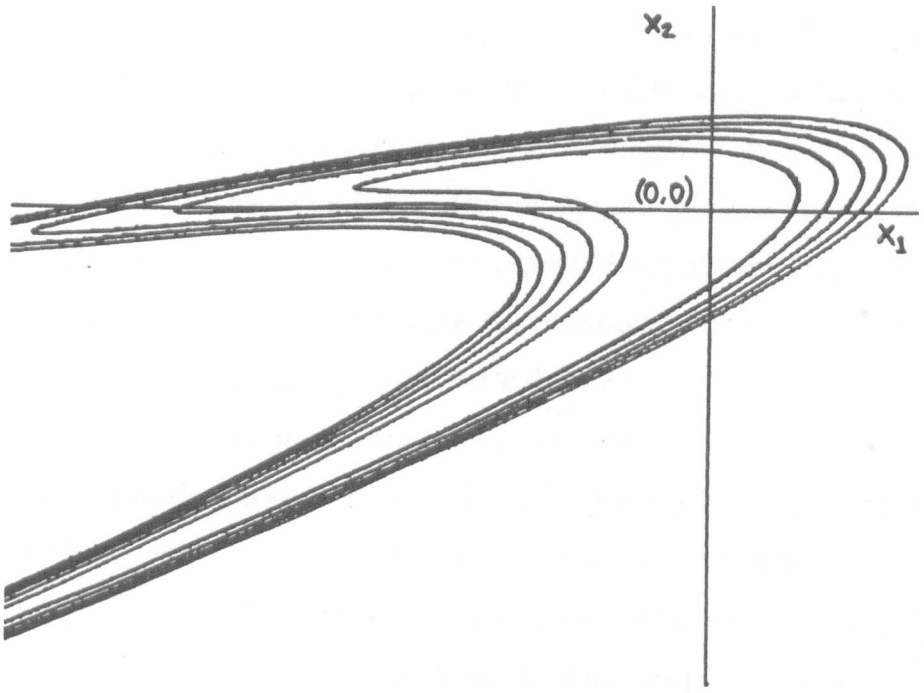


Figure 2.4 Merit function contours for a non-linear system.

solution $(0,0)$ to the starting point $(10,-2)$. For a zero damping factor, the minimum is reached in a single step but as the damping factor is increased the applied parameter changes become smaller and the direction tends towards that of steepest descent. Hence the damping factor serves not only to limit the magnitude of the applied changes but also to alter the direction taken in parameter space. A steepest descent increment is used far from the solution continuously changing to a Gauss-Newton solution close to the minimum. The aberrations of optical systems are, however, non-linear functions of the system parameters and so the minimum cannot be reached in a single step. The effect of the non-linearities is to twist the hyper-ellipsoids which, combined with ill-conditioning, produces long and narrow twisting valleys in parameter space. This is illustrated in figure 2.4 which again shows the merit function contours for values $\Psi = 4, 8, 12, 16, 20$ for the non-linear system

$$f_1 = x_1 + 0.3x_2 + (0.3x_1 - x_2)^2 \tag{2.23}$$

$$f_2 = 0.25x_1 - 0.25x_2 \tag{2.24}$$

The method of damping in the above equations is known as 'additive damping'

because a positive constant p^2 is added to all terms on the diagonal of the matrix $\mathbf{A}^T\mathbf{A}$. Levenberg [2.16] also described a method where the functions which have high linear rates of change are more heavily damped than those with slow rates of change. The effect is to multiply the elements on the diagonal by $(1+p^2)$, and hence this method is usually termed 'multiplicative damping.' Meiron [2.22] has applied multiplicative damping to lens design and reported significant improvements in convergence in non-linear regions and when parameters of different types are used . However, Kidger and Wynne [2.19] found little difference between the two methods except when the parameter increments used for computing the finite difference differentials were poorly chosen. Jamieson [2.23] subsequently pointed out that there is little theoretical justification for either additive or multiplicative damping other than to limit the applied parameter changes in the presence of non-linearity. For example, damping the aberrations by an amount proportional to their linear rates of change is unjustified because it is the second order derivatives that limit the range of applicability of the linear approximation. The same argument led Buchelle [2.24] to introduce approximate second order damping where only the homogeneous second order derivatives were evaluated. Their values were then used as the damping factors for each aberration. For example, if the second order derivative is zero then the linear approximation allows the solution to be reached in a single step. However, larger values of the second order derivatives indicate a significant non-linearity and hence the applied parameter change should be reduced. The homogeneous second order derivatives have also been used by Huber [2.25] but for a different purpose which will be described in the next section.

In practical lens systems it is usually necessary to 'weight' the individual aberrations. The merit function is then defined by

$$\Psi(\mathbf{x}) = \sum_{i=1}^{M} w_i^2 f_i^2 \tag{2.25}$$

where w_i are the weighting factors for the aberrations f_i. The weighting factors in damped least squares perform two vital functions: they allow the designer to assign values of relative importance to the various aberrations and they also balance the effects of the varying rates of change and inherent differences in the magnitudes of

the different aberrations (wavefront aberrations, transverse ray aberrations, paraxial quantities). As a result, the final aberration balance is fundamentally affected by the choice of weighting factors. Most commercially available lens design programs which use the DLS optimization method have a set of default weighting factors which generally will need to be adjusted as a design is optimized.

All lens design optimization problems will have boundary conditions or other constraints, for example, lens edge thicknesses, paraxial quantities, real glasses etc. These can be handled in three possible ways: by freezing variables which violate a boundary, by treating the violation as another aberration targeted to the boundary value or by solving for the exact boundary value using Lagrangian multipliers. The relative merits of these methods has been discussed at length by several authors, most notably Kidger [2.26] and Hopkins and Kadkly [2.27].

The damped least squares algorithm was the first algorithm that met with success in optical design. It has remained the most widely used and developed optimization method. However, its major limitation is that it will only make changes where the merit function decreases. Consequently it can only find the local minimum solution to the starting design for a given set of aberration weights. This dependence on the starting design is of great importance and benefits from the input of an experienced designer. Another limitation of the damped least squares algorithm is that the final solution is dependent upon the choice of damping factor at each iteration and the selection of the aberration weighting factors. Both of these variables influence the direction taken in parameter space.

2.4 Enhanced Damped Least squares Algorithms

A majority of the enhancements to the damped least squares algorithm have been concerned with the rate of convergence to the local minimum. This was particularly important when the speed of even large timeshare mainframe computers was significantly slower than most modern personal computers.

In §2.2 a linear approximation was made to the aberration variation by truncating a Taylor series expansion after the linear term (equation (2.4)). The merit function, defined in equation (2.5), can therefore be written as

$$\Psi(x) = \sum_{i=1}^{m} \left(f_o + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} x_j \right)^2 \tag{2.26}$$

which includes quadratic terms of the form

$$\sum_{i=1}^{m} 2 \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial f_i}{\partial x_j} \frac{\partial f_i}{\partial x_k} x_j x_k \tag{2.27}$$

If the Taylor series expansion includes the second order terms then equation (2.26) becomes

$$\Psi(x) = \sum_{i=1}^{m} \left(f_o + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} x_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f_i}{\partial x_j \partial x_k} x_j x_k \right)^2 \tag{2.28}$$

which includes quadratic terms of the form

$$\sum_{i=1}^{m} \left(2 \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial f_i}{\partial x_j} \frac{\partial f_i}{\partial x_k} x_j x_k + f_0 \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f_i}{\partial x_j \partial x_k} x_j x_k \right) \tag{2.29}$$

for a system with m aberrations and n variables. Comparing expressions (2.27) and (2.29), the missing quadratic terms are the second order partial derivatives of the aberrations with respect to the variables. This was first noted by Grey [2.28] and also by Wynne and Wormell [2.29]. Dilworth [2.30] modified the basic damped least squares algorithm by taking account of the missing second order derivatives. A complete set of both homogeneous and inhomogeneous derivatives was considered too expensive to calculate and so only the homogeneous second order derivatives

were approximated. The method was termed "pseudo-second-derivative" (PSD) for this reason. The inhomogeneous terms are in any case smaller than the homogeneous terms for reasons noted by Wynne and Wormell [2.29].

Dilworth [2.30] reported faster convergence and an ability to avoid stagnation in the presence of differing non-linearities. Further examples of the application of the pseudo-second derivative algorithm were given in a later paper [2.31]. Faggiano [2.32] subsequently reported a further modification to the pseudo-second derivative method which involved selecting the smallest possible damping factor for which the merit function decreases. This was claimed to enhance the speed of convergence yet further.

Robb [2.33] also described a method for accelerating the convergence of the basic DLS algorithm which was based on the observation that when the optimization is progressing slowly, the direction of the applied solution vector does not change significantly. When this condition is detected, the solution vector can be scaled and reapplied without calculating and resolving the matrix equation. The resulting lens parameters are therefore given by

$$x_{n+1} = x_n + \lambda^n x^* \tag{2.30}$$

where x^* is the calculated solution vector and λ a suitably chosen scalar. This procedure is repeated until the merit function increases or the value of n exceeds a specified number of iterations. A convergence improvement factor of ≈ 11 was reported.

The enhancement to the DLS algorithm reported by Huber [2.34] [2.25] was concerned with approximating the homogeneous second order derivatives by the expression

$$\frac{\partial^2 f_i}{\partial x_j^2} \approx \frac{\frac{\partial f_i}{\partial x_j}' - \frac{\partial f_i}{\partial x_j}}{\delta x_j} \tag{2.31}$$

where the first-order derivatives had been calculated from the two previous least squares iterations and the change applied to the variable x_j was δx_j between these

iterations. Huber recognised that the major expense during optimization is the calculation of the optical change matrix. However, a knowledge of the second order derivatives allows the first-order change matrix to be updated without any extensive raytracing of the system by using the expression

$$\frac{\partial f_i{}'}{\partial x_j} = \frac{\partial f_i}{\partial x_j} + \frac{\partial^2 f_i}{\partial x_j^2}\delta x_j. \tag{2.32}$$

This equation performs an extrapolation from the second order derivative and hence Huber called his technique 'extrapolated least squares' (ELS). The updating of the first-order derivatives by this method can continue until the approximation breaks down due to higher order non-linearities thus saving a large amount of computing time.

Other techniques for enhancing the basic DLS algorithm have included different algorithms for solving the DLS equation (2.18), for example, Cholesky decomposition to form the transformation to orthogonal variables [2.17] and singular value decomposition [2.35]. Analytical derivatives have also been successfully employed to improve the accuracy and reduce computing time [2.36]. Many improvements that have not been documented will have also been made by working optical designers; therefore this cannot be considered to be an exhaustive survey.

2.5 Orthonormalization

Grey developed the method of orthonormalization [2.28] because of two particular areas where design with Hamiltonian aberration coefficients is intractable. Firstly, it is possible for two classical aberration coefficients to have finite values which compensate each other; the system is therefore at an optimum solution although a merit function defined as the sum of the squares of the aberration coefficients (equation (2.5)) would not be zero. Secondly, in general, each system parameter affects every aberration coefficient. Thus a parameter change that improves one

aberration coefficient may cause the values of other aberration coefficients to become significantly worse.

Orthonormalization counters both these problems by defining a set of aberrations such that

$$\frac{\partial}{\partial g_i}\frac{\partial P}{\partial g_j} = 0, \quad (i \neq j) \tag{2.33}$$

where P is a merit function and g_i the orthogonal aberrations. A consequence of equation (2.33) is that any decrease in an aberration will also result in a decrease in the merit function. The orthogonal aberration coefficients can therefore be targeted to zero which generally results in a satisfactory balance of the classical aberration coefficients. A set of parameters u_j, must also be defined such that

$$\frac{\partial g_i}{\partial u_j} = 0 \quad (i \neq j) \tag{2.34}$$

and therefore a change in one parameter will only affect one orthogonal aberration. By Grey's definition, an "aberration theory" is orthonormal if it is orthogonal (equation (2.33)), has an orthogonal parameter base (equation (2.34)) and if

$$\frac{\partial g_i}{\partial u_i} = 1. \tag{2.35}$$

The consequences of an orthonormal aberration theory can be most simply illustrated by examining the merit function contours in a fashion analogous to the previous section for the damped least squares algorithm. Figure 2.5 shows the merit function levels at values of $P = 4, 8, 12, 16, 20$ for the orthonormal linear system

$$g_1 = 1u_1 + 0u_2 \tag{2.36}$$

$$g_2 = 0u_1 + 1u_2 \tag{2.37}$$

which consists of two aberrations, g_1 and g_2, which are orthonormal with respect to the two variables u_1 and u_2. The merit function surfaces have been transformed in to hyperspheres and any solution of the least squares equations is directed towards the minimum $(0, 0)$.

The implementation of this method as a computer program has been discussed by Grey ([2.37] and [2.2]). The mathematical technique for forming the orthogonal

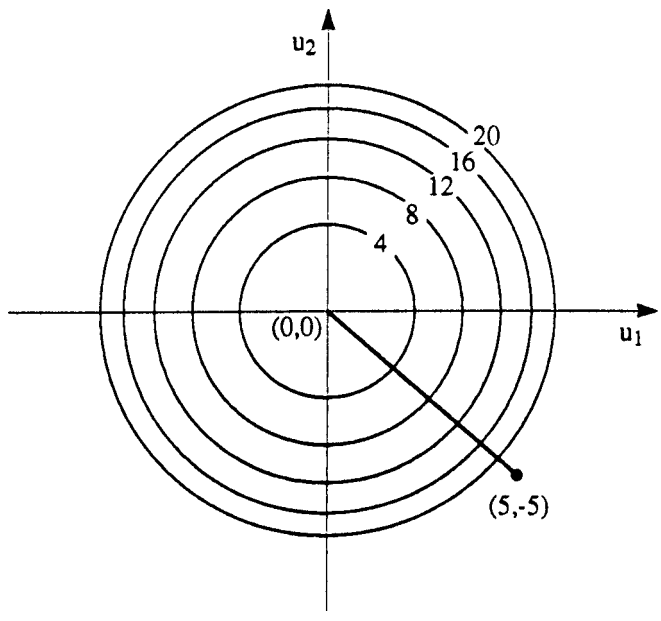


Figure 2.5 Contours of constant merit function for an orthonormal system.

parameter base is the Classic Gram-Schmidt process, a description of which can be found in Jamieson [2.23] or a number of mathematical texts (for example [2.38]). It is a linear transformation method and, since optical systems are non-linear, the orthogonal parameter base must be recomputed once the region of linearity has been exceeded. However, Grey noted that not all of the change matrix **A** is required for the computation of each orthogonal parameter and therefore it is possible to postpone the calculation of the relevant columns of **A** until after a preceding orthogonal change has been made [2.39]. This increases the capability of the orthonormalization method in the presence of non-linearity, a fact confirmed by the results of Takahashi [2.1].

The classic Gram-Schmidt orthogonalization method has numerical difficulties in its implementation which have been extensively discussed by Grey ([2.2] & [2.37]). However, these can be alleviated by using the modified Gram-Schmidt scheme and this approach has been described together with the results of some numerical work by Cornwell *et al.* [2.39].

2.6 The Adaptive Method

The adaptive method, first described by Glatzel [2.40], represents a significantly different approach to the problems of automatic correction of optical systems. In the adaptive method there must be at least as many variable parameters as there are aberrations during an iteration. This is the inverse situation to least squares where there must always be more aberrations than variables. Glatzel's reasoning for such an approach was that it is not possible to correct more *independent* aberrations than there are available variable parameters. The classical aberrations are not orthogonal and hence there are dependencies of various degrees between them. Glatzel's method exploits this by detecting linear dependence and removing one of the aberrations causing the interdependence. The number of aberrations is therefore restricted and it becomes practical to correct lens systems while maintaining at least as many variables as aberrations.

A linear system with more variables than unknowns is said to be underdetermined i.e. there is no unique solution. One approach to solving the system is to calculate the solution for the smallest change in the variables. This is known as a minimum norm solution since it minimizes the length of the solution vector $\mathbf{x}$. It is given by [2.41]

$$\mathbf{x} = \mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{A}^{\mathbf{T}})^{-1}(\mathbf{f_t} - \mathbf{f_0}) \tag{2.38}$$

where $\mathbf{x}$ is the vector of parameter changes with components $(x_1, x_2, \ldots, x_j, \ldots, x_n)$, $\mathbf{A}$ is the $m \times n$ matrix of first-order partial derivatives with elements $a_{ij} = \partial f_i / \partial x_j$, $\mathbf{A}^{\mathbf{T}}$ is its transpose and $(\mathbf{f_t} - \mathbf{f_0})$ is the m-vector of requested aberration changes.

Glatzel was not the first worker to use equation (2.38) to find the specified target values for the aberrations of an optical system. Hopkins and his co-workers at Rochester [2.42] first described a program for correcting third order aberrations for an exactly determined linear system. MaCarthy [2.43] then extended this to the minimum norm solution for situations when there were more degrees of freedom in the lens system. However, although equation (2.38) is the mathematical basis of

the adaptive method, it is really the control of the aberrations and the handling of the parameter changes that is significantly different.

In the adaptive method [2.44], the designer selects a small number of aberrations which are to be targeted to specified values and the lens parameters that the program may change. In addition a set of aberrations can be defined which are controlled to be within a specified tolerance band. Controlled aberrations do not take part in the correction process unless they have drifted outside their predetermined limits. To counter non-linearity, the target values are subdivided into a series of intermediate levels by dividing the requested aberration changes by a fixed factor entered in the program. Equation (2.38) is then solved using the first and subsequent intermediate target levels for the requested parameter changes $\mathbf{x}$.

The parameter changes are not all applied to the system simultaneously but are first divided by a scalar constant and then the largest element of $\mathbf{x}$ is applied to the system. The reason for this is that the higher order aberrations change relatively slowly and hence have the strongest influence on the direction taken in parameter space. Therefore, applying the largest element of $\mathbf{x}$ first should preferentially reduce the higher order aberrations. After applying the largest requested single variable change, the system is evaluated and the process continues until the region of linearity has been exceeded. The largest factor may be the same variable for several cycles but may then switch to another variable. This process of applying several parameter increments without recomputing the change matrix $\mathbf{A}$ was termed "inner cycles" by Glatzel and it is one of the reasons for the rapid convergence of the method. Figure 2.6 illustrates this process for a single non-linear aberration and just one variable. In practice, the function f_1 would be the vector $\mathbf{f}$ containing all aberrations and the variable x_1 could change as described earlier in this paragraph.

During correction the matrix $\mathbf{A}$ is tested for ill-conditioning. This occurs when two rows or columns become nearly linearly dependent and is countered by removing one of the interdependent aberrations. It is also possible for variables to

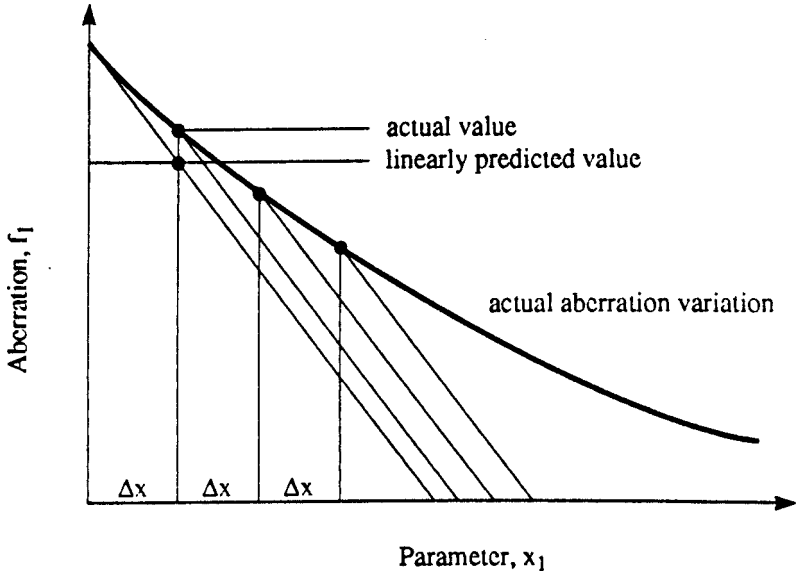


Figure 2.6 Illustration of 'inner cycles' in Glatzel's adaptive method.

violate boundaries, in which case they are frozen and removed from the vector **x** of variable parameters. Whenever the number of aberrations or variables changes, equation (2.38) is resolved; it only becomes necessary to recompute the change matrix when the linear approximation breaks down, or when the current target levels have been reached. The optimization stops when either all aberrations have reached the requested target values or when the number of variables becomes less than the number of aberrations.

At about the same time as Glatzel first published his method, Hopkins and Spencer [2.45] and Spencer [2.46] published their development of an optimization program which was along similar lines to the adaptive method. However the success of other optimization programs at the time, notably the damped least squares algorithm, meant that both versions of the adaptive method went almost unnoticed. As a result they have subsequently received little attention in the literature [2.47–2.49]. However the adaptive method has some important and useful features which have formed the basis of some of the work elsewhere in this thesis (see Chapter 3).

2.7 Global Optimization

The techniques presented earlier in this chapter are all local minimization methods; they can only take steps which decrease the merit function and hence the solution will be the local optimum for the starting design. (An exception to this is the adaptive method which, although constrained to always accept decreases in the targeted aberrations, is not constrained by the topography of a merit function surface. This will be discussed further in Chapter 3). The possibility, therefore, always arises that there could be a significantly better design at some other point in parameter space. The very best design possible with respect to the merit function specification is generally referred to as the global minimum although it is important to remember that there is no mathematical proof that any such solution is the global minimum.

2.7.1 Simulated Annealing

The simulated annealing algorithm attempts to model a molten metal as it is cooled. Initially, the atoms in the metal crystal have no preferred direction of orientation resulting in a high total energy state. If the metal is cooled slowly the atoms can align themselves along the preferred directions in the crystal lattice and the resultant total energy is at its global minimum. This process is known as annealing. However, if the metal is cooled too quickly the atoms cannot orientate themselves along their preferred directions and the resulting energy state is high. The simulated annealing algorithm can be applied to optical design if the energy of the crystal lattice is replaced by a merit function for the optical system being optimized.

The generalized simulated annealing algorithm proceeds as follows. Firstly, a random set of changes is generated in the construction parameters and this change is applied to the system. Secondly the merit function is evaluated and if it is lower than the previous value, the change is accepted with a probability of 1. However, if the merit function is greater than its predecessor a random number R is generated

in the range $[0, 1]$. The probability P is then evaluated from the expression

$$P = e^{-\beta(\Delta\Psi/\Psi)} \tag{2.40}$$

and if $P < R$ then the change is accepted but if $P > R$ then the change is rejected. In equation (2.40), β is a scaling factor, $\Delta\Psi$ the change in the merit function and Ψ the current value of the merit function. If β is set equal to $1/\kappa T$ then P becomes the Boltzman distribution and the analogy with cooling molten solids becomes more apparent. As a result, control of the parameter β is known as a 'cooling schedule.' In the basic algorithm, random changes are generated in the construction parameters until no further reduction in Ψ can be achieved. The system is then said to have reached equilibrium for the particular value of β. It is only at this stage that the value of β is incremented. The division by Ψ in the exponent of equation (2.40) ensures that uphill steps are accepted with a greater probability when the system is a long way from the global minimum. As the global minimum is approached, the chance of the algorithm accepting a change for which there is an increase in Ψ decreases thus speeding convergence to the solution.

The first reported application of simulated annealing to lens design was by Bohachesky *et al.* [2.50] at Los Alamos where it was to be used as global optimizer within an 'intelligent' optical design program. Further developments were given by Viswanathan *et al.* [2.51] although, to the authors' knowledge, no subsequent publications presented the results of their findings or practical examples from a functioning program. Their implementation of an 'intelligent' optical design program will be considered in the next section. Application of the generalized simulated annealing algorithm has also been reported by Hearn [2.52, 2.53] and Weller [2.54] who both describe implementations which use the damped least squares algorithm. The observation was that damped least squares converges very rapidly in the region of a solution, whereas simulated annealing finds good solution spaces, but then converges slowly. In addition, an initial run with a damped least squares algorithm can help limit the area of interest in the design space and information can be gained to help set up the simulated annealing variables.

In a later paper by Hearn [2.55], where simulated annealing was applied to cost control in optical systems, a modification to the probability function (equation (2.40)) was employed which was derived from the laws of statistical mechanics. The result was a continuous function which did not always accept the system changes if the merit function had decreased. It was given by

$$P = \frac{1}{1 + e^{-\beta(\Delta\Psi/\Psi)}} \tag{2.41}$$

and circumvented the problem that in the original algorithm (equation (2.40)) steps for which there was very little change in the merit function had an acceptance probability $P \approx 1$. Equation (2.41) was considered more 'natural' since very small values of $\Delta\Psi$ only had a 50% probability of acceptance.

More recently, Forbes and Jones [2.56] have been investigating adaptive simulated annealing for lens optimization. This approach rejects the original 'staircase' temperature schedule for β and recognizes that, at every stage of the simulated annealing process, there is an optimum step and associated temperature if the algorithm is to reach equilibrium rapidly. The method chosen, which is based on a heuristic principle, minimizes the distance to equilibrium at each cycle. Forbes and Jones [2.56] showed that, if this was the case, the temperature change depends on: the sensitivity of the equilibrium distribution, the reduction towards equilibrium and a constant entered to the program which determines the extent to which the parameter space is explored. As a result, the change in the temperature is modified at each cycle and hence the algorithm is termed 'adaptive simulated annealing'.

2.7.2 Other Global Search Techniques

Other approaches to global optimization have included a parametric survey technique where the design space is sampled with a regularly spaced grid and the best regions recorded for local minimization algorithms [2.57]. Such an approach is computationally very expensive since, in common with all global optimization algorithms, it suffers from the 'curse of dimensionality.' This is a simple pertinent

observation which states that if parameter space is to be sampled with I intervals for each parameter, then the number of system evaluations required is I^n where n is the number of variables. It is perhaps not suprising that results using global optimization techniques have typically been limited to 2 or 3 element designs.

Just recently Code V released a version of their optical design program with a "global optimizer." Although some initial results of its application have been presented by Kuper and Harris [2.58], no details of the actual algorithm are available at present.

2.8 Artificial Intelligence

Several papers have been published over the last seven years on the application of artificial intelligence to lens design [2.59]–[2.61], [2.64]. However, only Dilworth [2.61] has demonstrated a successful application of artificial intelligence that has been implemented in commercially available software (SYNOPSYS). His approach to the problem is twofold: a natural language interface is used to remove the problem of the designer having to remember the large number of commands necessary to control a comprehensive software package, and an expert systems approach is used to aid the design process. The former is concerned with clever data manipulation - the user can request the program to display data in many different combinations and forms. An example which illustrates the value of such an interface is that the user can request the program to plot coma versus the shape of a lens and display the output graphically.

In Dilworth's approach to 'expert systems', a database of lenses designed by experts is maintained. During correction of the aberrations of a particular design, the program considers splitting the lens at any air or cemented interface. The wavefronts are computed at these locations and the database is then searched for a lens or part of a lens, which has a wavefront that balances the wavefront of the aberrated lens,

the result obviously being a reasonably well corrected system. The design philosophy is to use the expert system to suggest lens designs which meet a preliminary specification that can then be subjected to further design work. The expert system therefore has the ability to provide input to an optimization routine, but does not deal with the fine correction of lenses at the optimization stage. This approach can be programmed in a 'standard' language, such as FORTRAN, and details of the procedures followed are given in the paper together with an example.

The expert systems approach to preliminary design has also been investigated by Moss [2.62]. Few commercial programs help the designer through the initial stages of the design process and therefore this has proved a valuable area for expert systems. The optical knowledge, design aids (e.g. the Delano diagram [2.63]) and rules of thumb can be implemented in software which can then suggest lenses for further design work. Moss achieves this by matching specifications with generic lens types held in a selection matrix where the matching is governed by a rule set. The resulting thin lens layout is optimized and then a routine is used to thicken the lens before it is re-optimized. Although such a design tool is of practical use, the system is limited by the elements in the selection matrix and the rule set used. Also the lens, resulting from a 'standard' optimization routine, may still be poorly balanced and difficult to manufacture.

The approach taken by Choudry and Johnson [2.64], considers the overall process of design in relation to specification, analysis, design, manufacture and test. The result is a knowledge base consisting of both commercial codes (algorithms) and heuristics generated using knowledge engineering. An inference engine is used to search the knowledge. No successful implementations of this level of artificial intelligence have been demonstrated, although several other authors have described a similar approach [2.59] & [2.51].

2.9 References

[2.1] T. Takahashi, "An Experimental Analysis of Optimization Algorithms Using a Model Function," Optik, **35,** 101–115 (1972).

[2.2] D.S. Grey, "Recent Developments in Orthonormalization of Parameter Space," in Proc. Conf. *Lens Design with Large Computers,* Rochester, NY, USA, 5–8 July 1966, 27-1—27-22.

[2.3] G. Black, "On the Automatic Design of Optical Systems," Proc. Phys. Soc, B, **68,** 729–736 (1955).

[2.4] D. Nicholls, *Private Communication*, 1989.

[2.5] A. Cauchy, Comptes Rendus, **25**, 536–538 (1847).

[2.6] D.P. Feder, "Automatic Lens Design Methods," JOSA, **47**, 902–912 (1957).

[2.7] C.G. Wynne, "Lens Designing by Electronic Digital Computer I," Proc. Phys. Soc., **73**, 777–787 (1959).

[2.8] J. Meiron and H.M. Loebenstein, "Automatic Correction of Residual Aberrations," JOSA, **47**, 1104–1109 (1957).

[2.9] D. Veerabhadra Rao and V. Venkateswara Rao, "Optimization Techniques in Lens Design," J. Sc. and Ind. Res. (India), **40**, 5, 297–312 (1981).

[2.10] D.P. Feder, "Automatic Optical Design," Appl. Opt., **2**, 1209–1226 (1963).

[2.11] J. Meiron and G. Volinez, "Parabolic Approximation for Automatic Lens Design," JOSA, **50**, 207–211 (1960).

[2.12] M.R. Hestenes and E. Stiefel, J. Res. Nat. Bur. Std., **49**, 409–436 (1952).

[2.13] D.P. Feder, "Automatic Lens Design with a High-speed Computer," JOSA, **52**, 172–176 (1962).

[2.14] S. Rosen and C. Eldert, "Least-squares Method for Optical Correction," J.O.S.A., **44**, 250–252 (1954)

[2.15] J. Meiron, "Automatic Lens Design by the Least-squares Method," J.O.S.A., **49**, 293–298 (1959)

[2.16] K. Levenberg, Quart. J. Appl. Math., **2**, 164 (1944).

[2.17] T.I. Harris and M.J. Hayford, "New Developments in Code V Optimization," in Proc. SPIE Vol. 237, 11–17, *International Lens Design Conference,* Oakland, CA, USA, 31 May - 4 June 1980.

[2.18] D.W. Marquardt, "An Algorithm for Least-squares Estimation of Nonlinear Parameters," J. SIAM, **11**, pp 431–441 (1963).

[2.19] M.J. Kidger and C.G. Wynne, "Experiments with Lens Optimization Procedures," Opt. Acta, **14**, 279–288 (1967)

[2.20] R. Fletcher, "A Modified Marquardt Subroutine for Non-linear Least-squares," A.E.R.E. (UK) Harwell Technical Report No. R6799 (1971).

[2.21] P.R. Adby and M.A.H. Dempster, *Introduction to Optimization Methods,* Chapman and Hall, London (1974), p 103.

[2.22] J. Meiron, "Damped Least-squares Method for Automatic Lens Design," J.O.S.A., **55,** 1105–1109 (1965)

[2.23] T.H. Jamieson, "Optimization Techniques in Lens Design," Adam Hilger Ltd., London (1971), p 19.

[2.24] D.R. Buchele, "Damping Factor for the Least-squares Method of Optical Design," Appl. Opt., **7,** 2433–2435 (1968)

[2.25] E.D. Huber, "Extrapolated Least-squares Optimization in Optical Design," J.O.S.A. A, **2,** 544-554 (1985)

[2.26] M.J. Kidger, PhD Thesis, University of London (1971).

[2.27] H.H. Hopkins and A. Kadkly, "Control of Damping Factors and Boundary Conditions in Optimization Programs," J. Mod. Opt., **35**, pp49–74 (1988).

[2.28] D.S. Grey, "Aberration Theories for Semi-automatic Lens Design by Electronic Computers I. Preliminary Remarks," JOSA, **53**, 672–676 (1963).

[2.29] C.G. Wynne and P.M.J.H. Wormell, "Lens Design by Computer," Appl. Opt., **2,** 1233–1238 (1963).

[2.30] D.C. Dilworth, "Pseudo-second Derivative Matrix and its Application to Lens Design," Appl. Opt., **17,** 3372–3375 (1978)

[2.31] D.C. Dilworth, "Automatic Lens Optimization: recent improvements," in Proc. SPIE Vol. 554, 191–196, *International Lens Design Conference,* Cherry Hill, NJ, USA, 10–13 June 1985.

[2.32] A. Faggiano, "Automatic Lens Design with Pseudo-second-derivative Matrix. A Contribution." Appl. Opt., **19,** 4226–4229 (1980).

[2.33] P.N. Robb, "Accelerating Convergence in Automatic Lens Design," Appl. Opt., **18,** 4191–4194 (1979)

[2.34] E.D. Huber, "Extrapolated least squares optimization: a new approach to least squares optimization in optical design," App. Op. **21**, pp 1705–1707 (1982).

[2.35] B. Tatian, "Nonlinearity and Lens Design", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, pp154–164 (1990).

[2.36] J. MacDonald and Y. Wang, "Analytical Derivatives in Damped Least-squares Optimization," in Proc. SPIE Vol. 655, 49–53, *Optical System Design, Analysis and Production for Advanced Technology Systems,* Innsbruck, Austria, 15–17 April 1986.

[2.37] D.S. Grey, "Aberration Theories for Semi-automatic Lens Design by Electronic Computers II. A Specific Program," JOSA, **53**, 677–680 (1963).

[2.38] J. Dunning-Davies, *Mathematical Methods for Mathematicians, Physical Scientists and Engineers,* Ellis-Horwood (1982), pp 186–187.

[2.39] L.W. Cornwell *et al.*, "Grey's Method for Non-linear Optimization," J.O.S.A., **63,** 576–581 (1973)

[2.40] E. Glatzel, "Ein neues Verfahren zur automatischen Korrektion optischer Systeme mit elektronischen Rechenmaschinen," Optik, **18**, 577–580 (1961)

[2.41] A.L. Peressini *et al.*, *The Mathematics of Non-Linear Programming*, Springer-Verlag, New York (1980) 145–148.

[2.42] R.E. Hopkins *et al.*, "Automatic Correction of Third-Order Aberrations ", J.O.S.A., **45**, 363–365 (1955).

[2.43] C. McCarthy, "A Note on the Correction of Third-order Aberrations", J.O.S.A., **45**, 1087–1088 (1955).

[2.44] E. Glatzel and R. Wilson, "Adaptive Automatic Correction in Optical Design," Appl. Opt., **7**, 265–276 (1968)

[2.45] R.E. Hopkins and G.H. Spencer "Creative Thinking and Computing Machines in Optical Design," J.O.S.A., **52**, 172–176 (1962)

[2.46] G.H. Spencer "A Flexible Automatic Lens Correction Procedure," Appl. Opt., **2**, 1257–1264 (1963)

[2.47] M. Krautter, "Zusammenhänge der in Programmen zur automatischen Kerrektion verwendeten Lösungmatrizen," Optik, **30**, 332–337 (1970)

[2.48] J.L. Rayces, "Ten Years of Design with Glatzel's Adaptive Method," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, pp75–84 (1980)

[2.49] W. Bessematter, "A Sophisticated Optical Design Program for Microcomputers," in Proc. SPIE Vol. **399**, *Optical System Design, Analysis and Production*, pp66–73 (1983)

[2.50] I. Bohachevsky, V.K. Viswanathan and G. Woodfin, "An 'Intelligent' Optical Design Program," in Proc. SPIE Vol. **485**, *Applications of Artificial Intelligence*, pp104–112 (1984)

[2.51] V.K. Viswanathan, I.O. Bohachevsky and T.P. Cotter, "An Attempt to Develop an 'Intelligent' Optical Design Program," in Proc. SPIE Vol. **554**, *International Lens Design Conference*, pp10–17 (1985)

[2.52] G.K. Hearn, " Generalized Simulated Annealing Optimization Used in Conjunction with Damped Least-squares Techniques" in Proc. SPIE Vol. **766**, *Recent Trends in Optical System Design*, pp283–284 (1987)

[2.53] G.K. Hearn, "Design Optimization Using Generalized Simulated Annealing," in Proc. SPIE Vol. **818**, *Current Developments in Optical Engineering II*, pp258–264 (1987)

[2.54] S.W. Weller, "Simulated Annealing: What good is it?" in Proc. SPIE Vol. **818**, *Current Developments in Optical Engineering II*, pp265–274 (1987)

[2.55] G.K. Hearn, "Automation of Cost Control During Optimization," in Proc. SPIE Vol. **1049**, *?*, pp??–?? (1989)

[2.56] G.W. Forbes and A.E.W. Jones, "Towards Global Optimization with Adaptive Simulated Annealing", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, pp144–153 (1990).

[2.57] D. Sturlessi and D.C. O'Shea, "Future of Global Optimization in Optical Design", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, pp54–68 (1990).

[2.58] T.G. Kuper and T.I. Harris, "A New Look at Global Optimization for Optical Design," Photonics Spectra, January 1992.

[2.59] Rong-Seng Chang, "Analytical Lens Design by Microcomputer with Artificial Intelligence," Proc. SPIE Vol. 554, *International Lens Design Conference,* (1985).

[2.60] S.W. Weller, "Optics Toolbox: an intelligent relational database system for optical designers," in Proc. SPIE Vol. 679, *Current Developments in Optical Engineering and Diffraction Phenomena,* (1986).

[2.61] D.C. Dilworth, "Applications of Artificial Intelligence to Computer-Aided Lens Design," in Proc. SPIE Vol. 766, *Recent Trends in Optical System Design: Computer Lens Design Workshop,* (1987).

[2.62] G.H. Moss, "An Expert System for Preliminary Optical Design," in Proc. SPIE Vol. 1013, *Optical Design Methods, Applications and Large Optics,* (1988).

[2.63] E. Delano, "First-order Design and the y, $\bar{y}$ Diagram," Appl. Opt., **2**, 1251–1256 (1963).

[2.64] A. Choudry and R.B. Johnson, "Knowledge Based Environment for Optical System Design," in Proc. SPIE Vol. 805, *Optical Componenets and Systems,* (1987).

CHAPTER 3

A Quadratic Adaptive Method for Aberration Correction

3.1 Introduction

In most optimization algorithms, the aberration variation is approximated by a Taylor series expansion [3.1] about its current value. For aberration f_i, which has a current value of $(f_i)_0$, the new value for this aberration after applying a set of parameter changes $(x_1, x_2, \ldots, x_j, \ldots, x_n)$ is given by

$$f_i = (f_i)_0 + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} x_j + \ldots \tag{3.1}$$

In all correction methods employing a linear approximation, the series is truncated after the linear term. The first-order partial derivatives are usually calculated using finite differences although they can be evaluated analytically [3.2], [3.3]. The offset finite difference approximation for a first derivative is given by

$$\frac{\partial f_i}{\partial x_j} \approx \frac{f_i(x_j + \delta x_j) - f_i(x_j)}{\delta x_j} \tag{3.2}$$

which requires two system evaluations; an evaluation of the current system and an evaluation of the system perturbed by an amount δx_j. If greater accuracy is required, the central finite difference method can be used but at the expense of a further system evaluation,

$$\frac{\partial f_i}{\partial x_j} \approx \frac{f_i(x_j + \delta x_j) - f_i(x_j - \delta x_j)}{2\delta x_j}. \tag{3.3}$$

Equation (3.2) provides an adequate approximation under most conditions, although greater efficiency both in terms of increased accuracy and reduced raytracing load has been reported when using analytical derivatives in the damped-least squares algorithm [3.4]. If the second order terms are included in the Taylor series expansion, equation (3.1) becomes

$$f_i = (f_i)_0 + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} x_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} x_j x_k \frac{\partial^2 f_i}{\partial x_j \partial x_k} + \ldots \tag{3.4}$$

which can be rewritten in matrix notation as

$$f_i = (f_i)_0 + \mathbf{d} \cdot \mathbf{x} + \tfrac{1}{2} \mathbf{x} \cdot \mathbf{H} \cdot \mathbf{x} + \ldots \tag{3.5}$$

where $\mathbf{d}$ is the n-vector of first derivatives and $\mathbf{H}$ is the Hessian matrix of second derivatives. A finite difference approximation for homogeneous second derivatives is given by

$$\frac{\partial^2 f_i}{\partial x_j^2} \approx \frac{f_i(x_j + \delta x_j) - 2 f_i(x_j) + f_i(x_j - \delta x_j)}{\delta x_j^2}. \tag{3.6}$$

However, it is advantageous to calculate or approximate all terms in the Hessian matrix and not just the homogeneous terms. This is because in the region of a local minimum of a continuous function, a quadratic approximation allows the minimum to be found in one step. Even if the minimum is not sought, within a finite region of the starting point a quadratic approximation allows any value within that region to be found in a single step if the approximation is valid. For this reason, some linear minimization techniques build up information about the Hessian matrix over a series of iterations [3.5], [3.6]. However, if the aberration variation is approximated by equation (3.5), a function which has been specified by the user, then the Hessian matrix is known.

The quadratic correction method described here is an extension of the quadratic extrapolation algorithm given by Maxwell and Hull [3.7] operating in an environment derived from the adaptive method as described by Glatzel and Wilson [3.8]. The quadratic approximation allows large steps to be taken in parameter space, permits the explicit definition of intractable aberrations, and removes the repeated use of

the change matrix which was termed "inner cycles" by Glatzel and Wilson [3.8] thus greatly increasing efficiency.

3.2 Mathematical Basis of the Method

In lens correction methods that make direct use of aberration coefficients, it is usually necessary to *solve* a set of linear equations for the specified target values. A least squares solution is not suitable because it is only a *best compromise solution* to the requested aberration changes. However, in the former case, there must be at least as many variables as aberrations and the mathematical system is said to be underdetermined. This may appear to be an unnecessary restriction but, as was noted by Glatzel and Wilson [3.8], it is not possible to correct more *independent* aberrations than the number of available variable parameters.

For an underdetermined linear system, the minimum norm solution (the solution that minimises the length of the parameter change vector, $\|\mathbf{x}\|$) is given by [3.9]

$$\mathbf{x} = \mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{A}^{\mathbf{T}})^{-1}(\mathbf{f_t} - \mathbf{f_0}) \tag{3.7}$$

where $\mathbf{x}$ is the vector of parameter changes with components $(x_1, x_2, \ldots, x_j, \ldots, x_n)$, $\mathbf{A}$ is the $m \times n$ matrix of first-order partial derivatives with elements $a_{ij} = \partial f_i / \partial x_j$, $\mathbf{A}^{\mathbf{T}}$ is its transpose and $(\mathbf{f_t} - \mathbf{f_0})$ is the m-vector of requested aberration changes. The matrix product $(\mathbf{A}\mathbf{A}^{\mathbf{T}})^{-1}$ is one form of the generalised inverse of matrix $\mathbf{A}$. When there are an equal number of aberrations and variable parameters equation (3.7) reduces to

$$\mathbf{x} = \mathbf{A}^{-1}(\mathbf{f_t} - \mathbf{f_0}) \tag{3.8}$$

which is the well known matrix equation for solving an exactly determined system of linear equations.

Equations (3.7) and (3.8) above are based on a linear variation of the aberrations with respect to the variable parameters (equation (3.1)). The quadratic approximation to the aberration variation is an extension of the work of Maxwell and Hull [3.7] and it proceeds as follows.

Initially, equation (3.7) is solved to find the parameter change vector, $\mathbf{x}$, which is the set of changes required to reach the specified aberration targets on the assumption that the system is linear. The vector $\mathbf{x}$ is then mutiplied by a factor FS (the quadratic fitting step) and the resulting change vector applied to the system. This perturbed system is then evaluated and the new values of the aberrations $(f_i)_{FS}$ recorded. It is now possible to approximate the aberration variation, with respect to the parameter changes defined by the vector $\mathbf{x}$, with a quadratic function of the form

$$(f_i)_{s_i} = (a_0)_i + (a_1)_i s_i + (a_2)_i s_i^2 \tag{3.9}$$

where s_i is the length of the vector $\mathbf{x}$ for aberration f_i and the coefficients a_0, a_1 and a_2 have a subscript i to indicate that they refer to the quadratic approximation for aberration f_i (see fig. 3.1).

The coefficients $(a_0)_i$, $(a_1)_i$ and $(a_2)_i$ in equation (3.9) are determined by first considering the case when $s_i = 0$ which corresponds to the current or starting point before any change has been applied to the system. It follows immediately that

$$(a_0)_i = (f_i)_0 \tag{3.10}$$

where $(f_i)_0$ is the current value of aberration f_i. The first-order coefficient, $(a_1)_i$, is the gradient of aberration f_i with respect to s_i at $(f_i)_0$,

$$(a_1)_i = df_i/ds_i|_{(f_i)_0}. \tag{3.11}$$

It is calculated as follows: the change in the value of aberration f_i after applying a set of parameter changes $s_i\mathbf{x}$ is given by

$$\begin{aligned} \delta f_i &= \frac{\partial f_i}{\partial x_1} s_i x_1 + \frac{\partial f_i}{\partial x_2} s_i x_2 + \ldots + \frac{\partial f_i}{\partial x_n} s_i x_n \\ &= s_i \left\{ \frac{\partial f_i}{\partial x_1} x_1 + \frac{\partial f_i}{\partial x_2} x_2 + \ldots + \frac{\partial f_i}{\partial x_n} x_n \right\} \end{aligned} \tag{3.12}$$

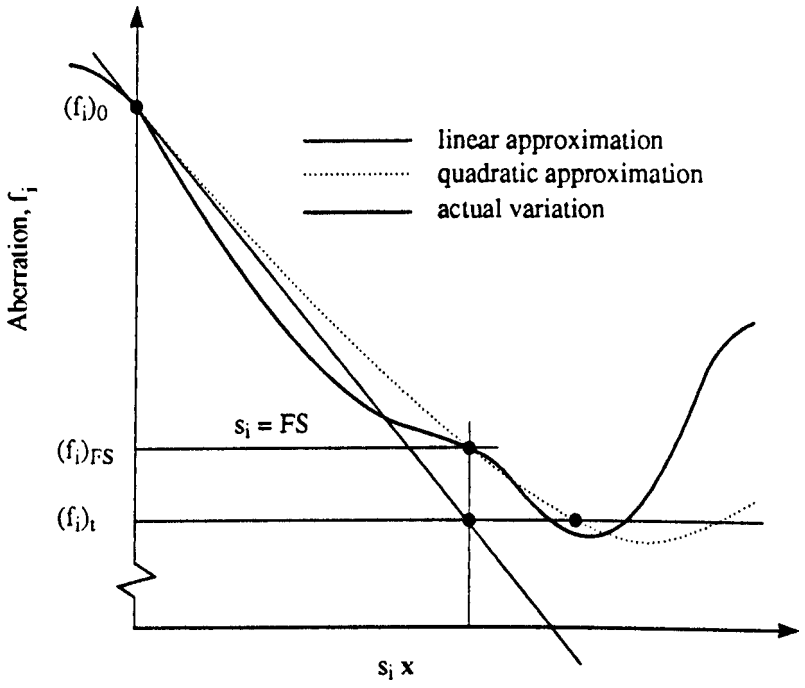


Figure 3.1. The relationship between the actual aberration variation and the linear and quadratic approximations.

on the assumption that the system is linear. When $s_i = 1$, the vector of parameter changes is simply $\mathbf{x}$ which is a solution of the underdetermined set of linear equations (3.7). Consequently, the aberration target values, $(f_i)_t$, will be reached and therefore, δf_i is simply the requested aberration change given by

$$\delta f_i = -((f_i)_0 - (f_i)_t) \tag{3.13}$$

Substituting equation (3.13) into equation (3.11) and recalling that $\delta s_i = 1$, gives

$$(a_1)_i = ((f_i)_t - (f_i)_0). \tag{3.14}$$

Finally the value of the aberration after applying a change $FS\mathbf{x}$ is $(f_i)_{FS}$ and hence substituting (3.10) and (3.14) into (3.9) and solving for a_2 gives

$$a_2 = \{(f_i)_{FS} + (FS - 1)(f_i)_0 - FS.(f_i)_t\} / FS^2. \tag{3.15}$$

The coefficients, $(a_0)_i$, $(a_1)_i$ and $(a_2)_i$, are calculated for each aberration and the quadratic approximations solved for the value of s_i which would achieve the specified aberration targets. Equation (3.9) will give two values for s_i and the larger is

discarded so as not to exceed the region of approximation. The values of s_i which are obtained for each aberration are converted in to a new set of aberration targets as follows.

According to a *linear approximation*, the variation of f_i with respect to s_i is given by

$$f_i = (f_i)_0 + a_1 s_i \tag{3.16}$$

and hence the *linear target value* which would achieve the specified aberration target *according to the quadratic approximation*, (the two values are not necessarily the same), is given by evaluating equation (3.16) for the value of s_i obtained from solving equation (3.9). If these target values are now used to form a new aberration change vector $(\mathbf{f_t} - \mathbf{f_0})'$ and equation (3.7) resolved for a new set of changes $\mathbf{x}'$, then the originally specified aberration targets will be reached *if the quadratic approximation is valid.*

The procedure described above, forms a single step or iteration during quadratic adaptive correction. Successive iterations are made until the target values are within the requested tolerances or no further correction is possible. The control of the aberrations, their target values and the variables which participate in each iteration, are described in the following section.

Finally, it is noted that the approximation in this method can be viewed as a smoothing function between the starting value of the aberration (point O) and the second fitting point (P in fig. 3.2). As such, only the Hessian matrix of this smoothing function is known. However, this will approximate to the Hessian of the actual variation to a greater or lesser extent depending on the goodness of fit, and hence this is a true second order method which approximates all second order derivatives, including the inhomogeneous terms.

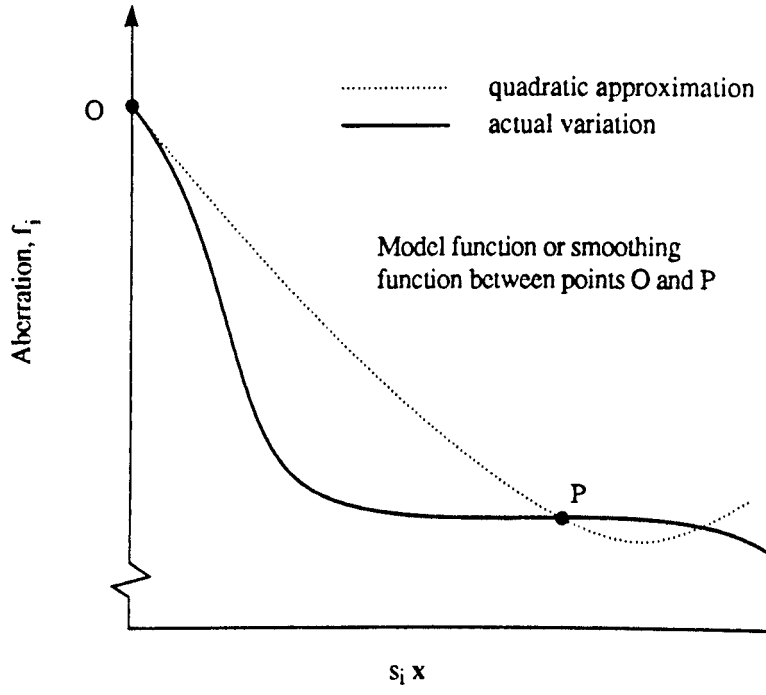


Figure 3.2. The quadratic approximation viewed as a smoothing function between points O and P.

3.3 Implementation of the Method

The quadratic correction method described in §3.2 is straightforward to program although there are several computational features which have not been described in the previous section. Some of these are to facilitate a flexible user-interface whereas other features are more fundamental to the success of the algorithm and it is these that will be described first.

3.3.1 Solving Equation (3.7)

Figure 3.3 shows a simplified flow-chart diagram of the computer program. First note that equation (3.7) is not solved in its stated form since the inversion of the matrix $(\mathbf{A}\mathbf{A}^{\mathbf{T}})$ introduces numerical errors and is rarely necessary when solving linear systems unless it is explicitly required. Instead, a unique vector $\mathbf{w}$ is defined such that $\mathbf{x} = \mathbf{A}^{\mathbf{T}}\mathbf{w}$ (Appendix B). Substituting this result into equation (3.7), gives

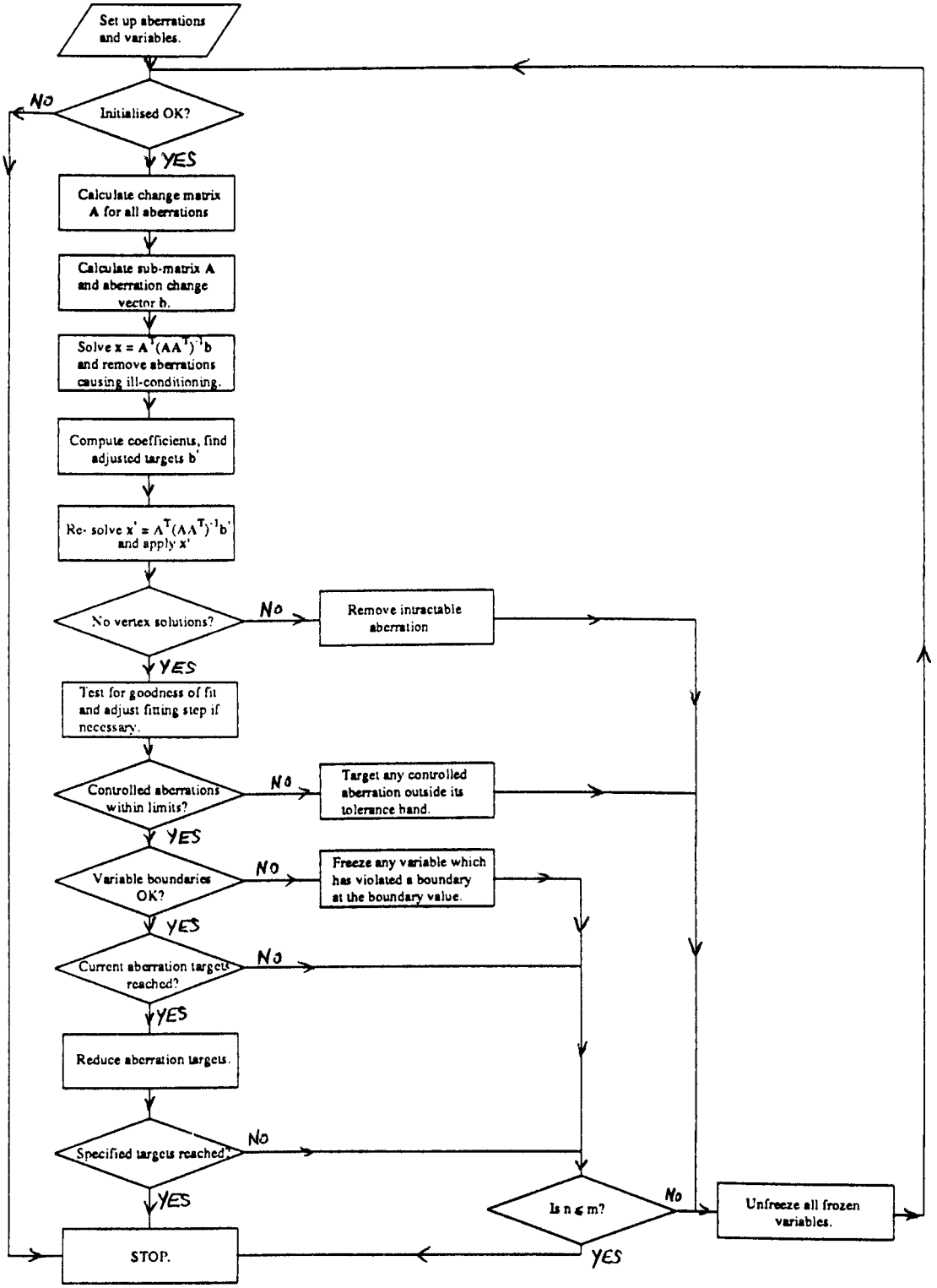


Figure 3.3. A simplified flow chart of the computer program for the adaptive quadratic correction algorithm.

$$\mathbf{A}^{\mathbf{T}}\mathbf{w} = \mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{A}^{\mathbf{T}})^{-1}(\mathbf{f_t} - \mathbf{f_0})$$

$$\mathbf{w} = (\mathbf{A}\mathbf{A}^{\mathbf{T}})^{-1}(\mathbf{f_t} - \mathbf{f_0})$$

$$(\mathbf{A}\mathbf{A}^{\mathbf{T}})\mathbf{w} = (\mathbf{f_t} - \mathbf{f_0}). \tag{3.16}$$

The matrix product $\mathbf{A}\mathbf{A}^{\mathbf{T}}$ is an $m \times m$ square matrix and hence (3.16) can be solved for $\mathbf{w}$ by LU decomposition [3.10] (other suitable methods that handle ill-conditioned equations well include QR decomposition [3.11] and singular value decomposition [3.12]). It is now possible to find $\mathbf{x}$ by evaluating

$$\mathbf{x} = \mathbf{A}^{\mathbf{T}}\mathbf{w}. \tag{3.17}$$

3.3.2 Ill-conditioning

The matrix $\mathbf{A}$ is well known for being ill-conditioned in optical systems. Ill-conditioning occurs when the matrix becomes near singular; that is when either two or more rows or columns are almost linearly dependent. The result of solving an ill-conditioned linear system is that a very large parameter change vector $\mathbf{x}$ is calculated which will not achieve the requested aberration target values. Usually this will slow the progress of the correction significantly, but in extreme cases stagnation can occur and no further improvement is possible until one of the offending aberrations is removed. It is therefore important to test for linear dependence between the rows of matrix $\mathbf{A}$ and remove one of the aberrations causing ill-conditioning. This helps to maintain rapid progress towards the solution.

An ill-conditioned linear system is therefore unstable with respect to very small perturbations in that system. For example consider the two linear equations

$$7x_1 + 10x_2 = 1 \tag{3.18}$$

$$5x_1 + 7x_2 = 0.7 \tag{3.19}$$

which have the solution $x_1 = 0, x_2 = 0.1$. However, if the right hand sides of equations (3.18) and (3.19) are perturbed to values of 1.01 and 0.69 respectively

then the solution becomes $x_1 = -0.17, x_2 = 0.22$. Therefore the changes in x are quite large compared to the perturbation 0.01. A measure of the ill-conditioning of a matrix is given by the ratio of the maximum eigenvalue to the minimum eigenvalue

$$cond(\mathbf{A}) = \frac{\text{Max}|\lambda|}{\text{Min}|\lambda|} \tag{3.20}$$

for the set of all eigenvalues on $\mathbf{A}$ [3.12]. One of the useful features of equation (3.20) is that it is independent of the matrix norm (the method used to determine the 'size' of the matrix). In optical problems the eigenvalues of the matrix products $\mathbf{A^T A}$ and $\mathbf{A A^T}$ are of interest since these appear in formulations of the damped least squares and the adaptive (minimum norm) algorithms respectively. The value of the condition number, $cond(\mathbf{A})$, has been calculated for several optical matrices and table 3.1 summarises the results. These change matrices have been calculated for a triplet lens where all six curvatures are variable. The matrix products are only useful for either the damped least squares or the adaptive method. For example, only examples 4–6 have a smaller number of aberrations than the variables and could be used to find a minimum norm solution in the adaptive method. As a result, values in the table have been omitted when the matrix products have no meaning to a practical correction method.

Aberrations in the matrix	$cond(\mathbf{A A^T})$	$cond(\mathbf{A^T A})$
1. All 7 primary aberrations	-	1.31×10^4
2. 7 primary aberrations & u'	-	6.58×10^7
3. Finite ray aberrations at 1 field position	-	2.34×10^7
4. S_I, S_{II}, S_{III}, C_I, C_{II}	2.27×10^4	-
5. s & t at 1 field position and u'	1.62×10^3	-
6. Finite TRA's at 1 off-axis field position	2.35×10^6	-

Table 3.1 The condition number, $cond(A)$, of the matrix products $\mathbf{A A^T}$ and $\mathbf{A^T A}$ for six optical change matrices

(Note: High condition number corresponds to ill-conditioning)

The values in table 3.1 are in good agreement with Feder [3.13], who has reported that values of $cond(\mathbf{A^T A}) = 10^6$ and higher can easily occur in optical systems. They also illustrate that the presence of aberrations in the change matrix, which

have widely differing rates of change, can significantly increase the degree of ill-conditioning. For example, compare examples 1 and 2 in table 3.1 where the introduction of the paraxial quantity u', which is used to control the effective focal length of the system, has increased the condition number by a factor of over 1000.

On their own the values in table 3.1 are not particularly meaningful because the question arises as to what value of $cond(\mathbf{A})$ can be tolerated before the instability created by ill-conditioning becomes significant. There is no easy answer to this question since it depends upon the accuracy of the numerical work and therefore upon computer hardware and software. As an indication, the condition number of the matrix, which must be inverted to solve equations (3.18) and (3.19), has been calculated and found to be $cond(\mathbf{A}) = 198$.

In most algorithms it is desirable to detect the presence of ill-conditioning so that the user can be warned of its presence and/or an offending aberration removed. In addition, the method should not add unduly to the computation time since its intended use would be during an optimization run. Two possible methods are the computation of the vector product and hence the angle between two row/column vectors or the computation of the ratios of elements in the two rows/columns being tested.

The linear angle between the two vectors $\mathbf{a}$ and $\mathbf{b}$ is given by the well known formula

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|} \tag{3.21}$$

and when the two vectors $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent such that $\mathbf{b} = r\mathbf{a}$ where r is a scalar, equation (3.21) becomes unity. Consider now the situation where the linear dependence is broken by the second element of $\mathbf{a}$ such that $\mathbf{a} = (a_1, fa_2, \ldots, a_n)$ where $f \neq 1$. In this case equation (3.21), after substituting $\mathbf{b} = r\mathbf{a}$ and expanding, becomes

$$\cos\theta = \frac{(a_1^2 + fa_2^2 + \ldots + a_n^2)}{\sqrt{(a_1^2 + f^2a_2^2 + \ldots + a_n^2)}\sqrt{(a_1^2 + a_2^2 + \ldots + a_n^2)}}. \tag{3.22}$$

Squaring both sides of equation (3.22) it can be seen that the numerator and de-

nominator only vary in terms containing a_2^2. The numerator contains terms in a_2^2 of the form

$$2fa_2^2(a_1^2 + a_3^2 + \ldots + a_n^2) \tag{3.23}$$

and the denominator terms of the form

$$fa_2^2(a_1^2 + a_3^2 + \ldots + a_n^2). \tag{3.24}$$

The only disparity between the expressions (3.23) and (3.24) is the factor of 2 in the cross-terms $fa_2a_n^2$. For a fixed value of f, θ would therefore be expected to depend on the magnitude of a_2. This may not be desirable since optical change matrices usually contain aberrations with largely differing rates of change. Then it becomes difficult when trying to compare the degree of linear dependence between rows or columns with very different magnitudes.

The second possibility for assessing the degree of linear dependence between two rows/columns of an optical change matrix is to evaluate the ratios of the corresponding elements in the two vectors and to compute their standard deviation. Consider the same two vectors **a** and **b** which are linearly dependent such that $\mathbf{b} = r\mathbf{a}$ where r is a scalar constant and where the second element in **a** is given by fa_2. If $f = 1$ then the vectors are parallel but if $f \neq 1$ then the linear dependence is broken. The vector **r** of the ratios of the elements in **a** and **b** is therefore given by

$$\mathbf{r} = (1/r, f/r, \ldots, 1/r) \tag{3.25}$$

and hence the mean ratio is

$$\bar{r} = \frac{\sum_{i=1}^{n} r_i}{n} = \frac{1}{r} - \frac{1-f}{rn}. \tag{3.26}$$

The standard deviation of the ratios is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} = \sqrt{(n-1)}\left(\frac{f-1}{rn}\right) \tag{3.27}$$

which is independent of the magnitude of element a_2 but dependent on the ratio r. Close to linear dependence $\bar{r} \rightarrow 1/r$ and hence

$$\sigma/\bar{r} \approx \sqrt{(n-1)}\left(\frac{f-1}{n}\right) \tag{3.28}$$

which is independent of the ratio r.

In the following numerical examples, the values of θ and $\sigma/\bar{r}$ have been computed for two row vectors which are nearly linearly dependent:- In all the examples, the value of one element of **a**, indicated by an asterisk, has been multiplied by $f = 1.05$ representing a 5% shift away from linear dependence.

$$\left.\begin{matrix} 1 & 1 & 2 & 3 & 5.25^* & 1 \\ 5 & 5 & 10 & 15 & 25 & 5 \end{matrix}\right\} \quad \theta = 1.35581, \quad \sigma/\bar{r} = 0.01848 \qquad (3.29)$$

$$\left.\begin{matrix} 1 & 1 & 2 & 3 & 5.25^* & 1 \\ 500 & 500 & 1000 & 1500 & 2500 & 500 \end{matrix}\right\} \quad \theta = 1.35581, \quad \sigma/\bar{r} = 0.01848 \qquad (3.30)$$

$$\left.\begin{matrix} 1.05^* & 1 & 2 & 3 & 5 & 1 \\ 5 & 5 & 10 & 15 & 25 & 5 \end{matrix}\right\} \quad \theta = 0.44146, \quad \sigma/\bar{r} = 0.01848 \qquad (3.31)$$

$$\left.\begin{matrix} 1 & 1 & 2 & 3.15^* & 5 & 1 \\ 5 & 5 & 10 & 15 & 25 & 5 \end{matrix}\right\} \quad \theta = 1.17270, \quad \sigma/\bar{r} = 0.01848 \qquad (3.32)$$

The first two examples (3.29) and (3.30) confirm that both methods of assessing linear dependence are independent of the row ratio r. Examples (3.31) and (3.32) demonstrate that the angle θ depends on the values of the vector element which breaks the linear dependence.

In the quadratic adaptive method, if the value of $\sigma/\bar{r}$ during each iteration drops below a specified level (0.05 was been found to be suitable), one of the rows is temporarily removed. Ill-conditioning that exists between two aberrations at one stage of the correction process may not be present at a later stage. Therefore, the aberration(s) that have been temporarily removed are regularly tested, and replaced into the matrix if the linear dependence is no longer present.

3.3.3 Specification of Aberrations during Correction

An important feature of the adaptive method is the dynamic control of the matrix **A**. However, before describing the ways in which the program selects the aberrations and variables which go to form matrix **A**, the two types of aberration which can be specified by the user for correction must be defined: a *targeted* aberration is one that the designer wishes the program to change to a specified value, whereas a *controlled* aberration is allowed to drift within a certain range or tolerance band. As a result there are four types of aberrations during correction: targeted aberrations, targeted aberrations that have been controlled due to intractability, controlled aberrations which are within their tolerance band and controlled aberrations which have been temporarily targeted because they have drifted outside their tolerance band. Only the first and last types of aberration go to form matrix **A** for a particular iteration.

3.3.4 Intractable Aberrations

The handling of intractable aberrations is one of the advantages of using a quadratic approximation. A linear approximation will always intersect the aberration target value, whereas a quadratic approximation will give a vertex solution if the required level is not intersected. This indicates that one of the following conditions has arisen: either the aberration is intractable for the current choice of variables and aberrations in matrix **A**; or the approximation is poor (higher order non-linearities); or the variables do not permit the aberration variation to intersect the target level. In the program, vertex solutions are recorded and a targeted aberration removed if two successive vertex solutions are obtained. The goodness of fit is controlled by changing the value of the quadratic fitting step, FS (§3.3.7). However, if an aberration remains intractable, the choice of variables must be changed.

3.3.5 An Illustrative Example of Aberration Control

The features described in §3.3.2 – 3.3.4 have the ability to alter the aberrations participating in the correction and this can help the program to juggle the aberrations in its search for the specified target values. This occurs most frequently during the final stages of correction where several aberrations are at or close to their minimum values. Table 3.2 shows the course of correction over three iterations for the first three Seidel aberrations of a 50mm f/4 Cooke triplet. All six curvatures are variable, S_I, S_{II} and S_{III} have all been assigned target values of 0.00005, and the effective focal length of the system has been controlled via the paraxial axial marginal ray angle, u', to within $\pm 1\%$ of its specified target value (-0.125).

Aberration	Current	Result	Target
Iteration 1			
(u'	-0.12500	-0.11959	-0.12500)
S_I	0.01396	-0.00002	0.00005
S_{II}	0.00032	0.00006	0.00005
S_{III}	-0.00640	0.00005	0.00005
Controlled aberration violation - u' is now targeted			
Iteration 2			
u'	-0.11959	-0.12500	-0.12500
S_I	-0.00002	0.00075	0.00005
S_{II}	0.00006	0.00037	0.00005
S_{III}	0.00005	0.00024	0.00005
Controlled aberration within limits - u' is released			
Iteration 3			
(u'	-0.12500	-0.12437	-0.12500)
S_I	0.00075	0.00005	0.00005
S_{II}	0.00035	0.00005	0.00005
S_{III}	0.00024	0.00005	0.00005
All targeted aberrations are within tolerance			

Table 3.2. A simple correction illustrating the handling of controlled aberrations.

The first iteration makes the major changes required and the resulting aberration

values are all close to their specified targets. However, the value of u' has drifted outside its tolerance band and is therefore targeted for the second iteration. The second iteration now contains four aberrations; the three original targeted aberrations and u'. The aberration values, after the second iteration, demonstrate that u' has been returned to its specified value. Therefore, u' is removed from the correction. Finally, in iteration three, the three targeted aberrations are brought to their target values and the correction stops.

Although this is a trivial example, the precision with which the aberration targets have been found is not uncommon with the quadratic correction method. Two possible explanations for this are that the quadratic approximation is particularly good in the cases studied and that this technique encourages the designer to restrict the number of targeted aberrations, so assisting the efficiency of the correction algorithm.

3.3.6 Control of Variables

The handling of variables also allows a continual updating of matrix **A** as the correction proceeds, but in a more familiar way. The variables are assigned boundary values and, if a violation is requested, they are frozen at the boundary value and their respective column removed from the change matrix. It would also be possible to treat the variables as aberrations whenever they violate a boundary as is commonly done in many damped-least squares algorithms. However, although this provides a smoother boundary which can help prevent oscillations, it would be impractical in the present technique, where it is required that there are at least as many variables as aberrations.

3.3.7 The Quadratic Fitting Step FS

The automatic control of the quadratic fitting step FS has been mentioned in §3.3.4. It is an important parameter because it controls the goodness of fit of the quadratic approximation and hence the precision with which target values can be found. Although there must be an optimum value for this quantity during correction, it is not possible to predict what its value will be beforehand and so the program continually updates its value during correction. Successive iterations where the goodness of fit is better than 70% result in increases in the fitting step of 0.1. If the deviation from the quadratic approximation is greater than 30% , then the fitting step is reduced by 0.05 to a minimu level of $FS = 0.1$. (The goodness of fit is measured by the maximum deviation of a targeted aberration from the predicted value at each iteration). These values are somewhat arbitrary but have been found to work adequately.

3.3.8 Aberration Target Levels

In most situations it is not sensible to attempt large aberration changes in a single step, although obviously the quadratic adaptive method has the ability to cope with larger changes than is possible with linear methods. In fact the ability to make large changes has been found to be a useful feature and can help to avoid local minima [3.14]. This is discussed further in §3.5. If the changes requested are large, then the aberration change is split into a series of intermediate (or reduced) target levels. There are two possibilities for assigning the number of reduced target levels: either scale the length of the aberration change vector $(\mathbf{f_t} - \mathbf{f_0})$; or scale the length of the solution vector $\mathbf{x}$. The two cases are linked via the linear relationships expressed in equation (3.7) and therefore it would not be expected for there to be a significant difference between them and indeed both have been implemented successfully. The parameter change vector has been used, since the efficiency of the quadratic approximation can be more directly maximised without taking account of the rates of change of the aberrations. The number of reduced aberration levels

r is therefore given by the integer value of

$$r = \frac{\| \mathbf{x} \|}{\| \mathbf{x} \|_{max}} \tag{3.18}$$

where $\| \mathbf{x} \|_{max}$ is specified by the user. If an inappropriate value of $\| \mathbf{x} \|_{max}$ is specified, then there will be either too many or too few intermediate aberration target levels. If there are too many, successful steps will be made repeatedly at each iteration and the number of intermediate target levels is reduced. However, if the program is taking more than a fixed number of iterations (typically 3 to 5) before reducing the target levels, the number of levels is increased. This feature helps the program exploit the valid range of the quadratic approximation by making the largest possible parameter changes while still maintaining progress towards the solution.

3.4 Examples

3.4.1 Primary Aberration Correction of an f/4 Triplet

The first example is an $f/4$ triplet with an $EFL = 50mm$ and field of view ± 15 degrees (Appendix C, TL3.1), for which correction of all seven primary aberrations (five Seidel and two primary chromatic aberrations) has been attempted. The starting design consists of two plano-convex lenses placed symmetrically about a bi-concave lens and the stop placed at surface 4, the second surface of the negative lens. All six curvatures and the two air spaces were allowed to vary and the glass types were fixed at LAK9 for the outer positive elements and SF8 for the central negative element. Although the correction of all the primary aberrations is artificial and not a useful step towards achieving a finished design, it does allow the quadratic approximation to be investigated since the actual aberration variation can be rapidly calculated from a paraxial ray trace. Figure 3.4 shows four examples of the quadratic approximation which are taken from the correction of example 1. Each graph shows the linear and quadratic approximations together with the actual aberration variation plotted in

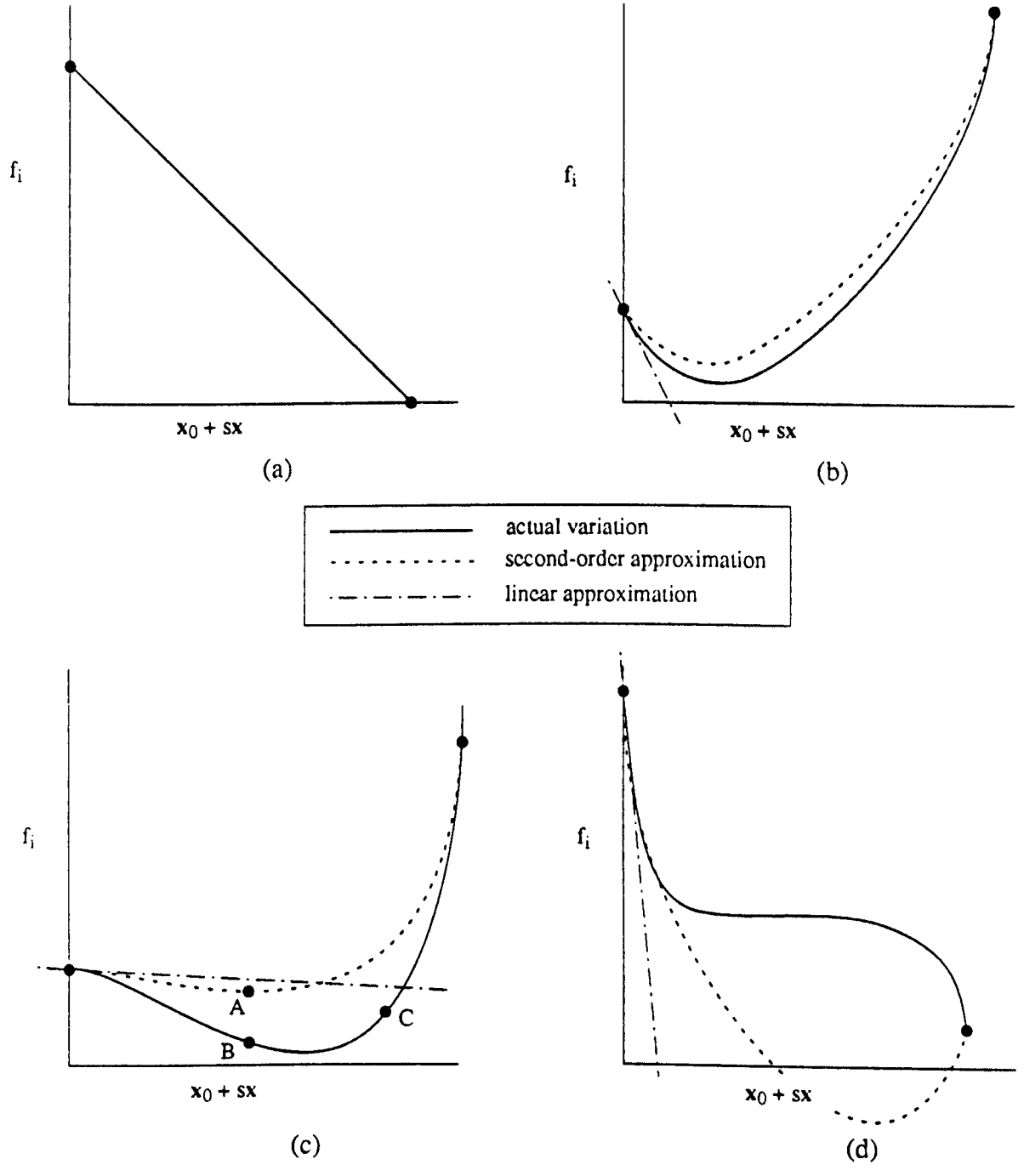


Figure 3.4 Actual aberration variations and approximations for (a) linear variation, (b) predominantly quadratic variation, (c) higher order variation where the QAC algorithm is successful and (d) higher order non-linearities requiring a smaller value of the quadratic fitting step FS.

the **x** direction. The actual variation has been found by evaluating the aberration for 100 points along the x-axis.

The following observations can be made from figure 3.4,When only small parameter changes are required, the algorithm is essentially linear but without the efficiency of the 'inner cycles' of the adaptive method (a). However, for a larger parameter change, the non-linearity is predominantly second order and the approximation is particularly good (b). Even when the initial approximation is poor, the QAC method can still find a valid approximation in a very few steps. For example, in (c) the aberration target value lies below the vertex of the quadratic approximation and hence the vertex solution will be taken (point A). If the system is evaluated and a new quadratic approximation computed from points B and C, the resulting approximation to the actual variation is particularly good. Finally, higher order non-linearities can cause significant problems (d) and in this case the situation can only be improved by reducing the value of the fitting step FS, giving a better approximation close to the starting value.

Example	Iterations	Steps
1. Primary correction of an f/4 triplet	14	2
2. Flat-field correction of an f/4 triplet	15	3
3. Final correction of an f/4 triplet	4	1
4. Correction of an f/1.7 50mm Double Gauss derivative	34	4

Table 3.3. The number of iterations for correction of examples 1–4 using quadratic adaptive correction.

Table 3.3 shows the number of iterations for the correction of this and other examples described below, together with the number of stages taken. The number of stages is recorded when the aberrations have been corrected sequentially. For example, in the correction of the $f/4$ triplet for all seven primary aberrations, the Petzval sum and distortion (S_{IV} and S_V) were corrected first since they require the largest changes in the system parameters. The correction of the five remaining

aberrations is achieved after S_{IV} and S_V have been reduced to a reasonable level. Simultaneous correction of all seven aberrations is not practical in this particular instance since ill-conditioning causes slow convergence. This technique is described further in §4.3.2.

3.4.2 Higher Order Astigmatism and Petzval Sum Balancing

In example 2, the higher order astigmatism and Petzval sum in the triplet of example 1 are balanced to achieve a flat image field. In addition, the axial spherical aberration is balanced. The aberration target values are given in Table 3.4 which is a typical example of the setup screen used in the program. All six curvatures are allowed to vary and the effective focal length is controlled by maintaining the paraxial axial marginal ray angle constant at its required value of -0.125.

Finite Ray Aberration	Current	Target	Tolerance(%)
dY' axial marginal ray	-0.16477	0.00500	50
t zonal-field principal ray	0.78512	0.03000	50
s full-field principal ray	-2.15128	0.03000	50
t full-field principal ray	1.62836	-0.09000	30

Table 3.4. Target values for the flat-field correction of an f/4 triplet (example 2).

The correction is achieved in 15 iterations (table 3.3) using the QAC algorithm and figure 3.5 shows the astigmatism curves for both the starting design and the corrected lens. The target values have been achieved although some of this correction may have to be balanced against off-axis aperture aberrations.

3.4.3 Further Correction of a 50mm f/4 Triplet

The third example presents a corrected design for the $f/4$ triplet and uses the flat-field solution of example 2 as its starting point. The only significant aberration that

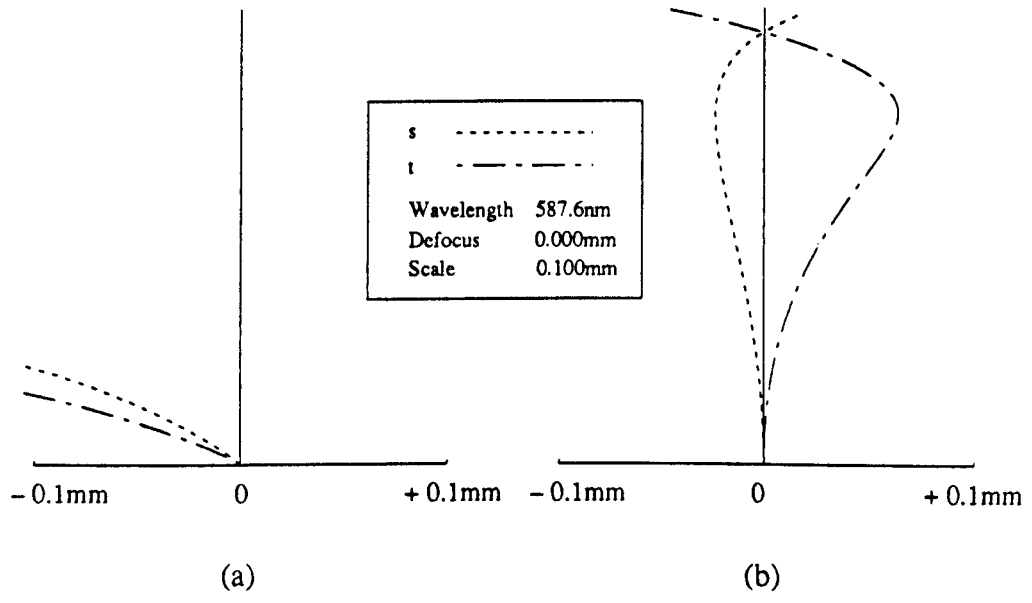


Figure 3.5. s & t Curves for an f/4 triplet (a) before and (b) after flat-field correction.

remains to be corrected is coma, which needs to be balanced against oblique spherical aberration. It is quite common that designs of the same general form as the triplet require some vignetting to achieve optimum performance and therefore the zonal aperture rays have been targeted. This allows the marginal rays to assume large aberration values since they can be vignetted at a later stage if necessary. During the course of correction, it was quickly discovered that only one off-axis field position could be corrected for coma if the flat-field correction was maintained. Attempts to correct all four zonal rays (two for each off-axis field position) produced a large number of intractable aberrations and the algorithm could make no improvement in the system. However, correction of each individual field position was possible and therefore it was necessary to balance the aberrational performance between the zone and the margin of the field. Figure 3.6 shows the resulting aberration curves achieved after 4 further iterations.

3.4.4 Correction of an f/1.7 50mm Double Gauss Lens

The triplet has only a small number of variables and so a double Gauss derivative has been chosen with an aperture of f/1.7, field of ±24 degrees and effective focal length

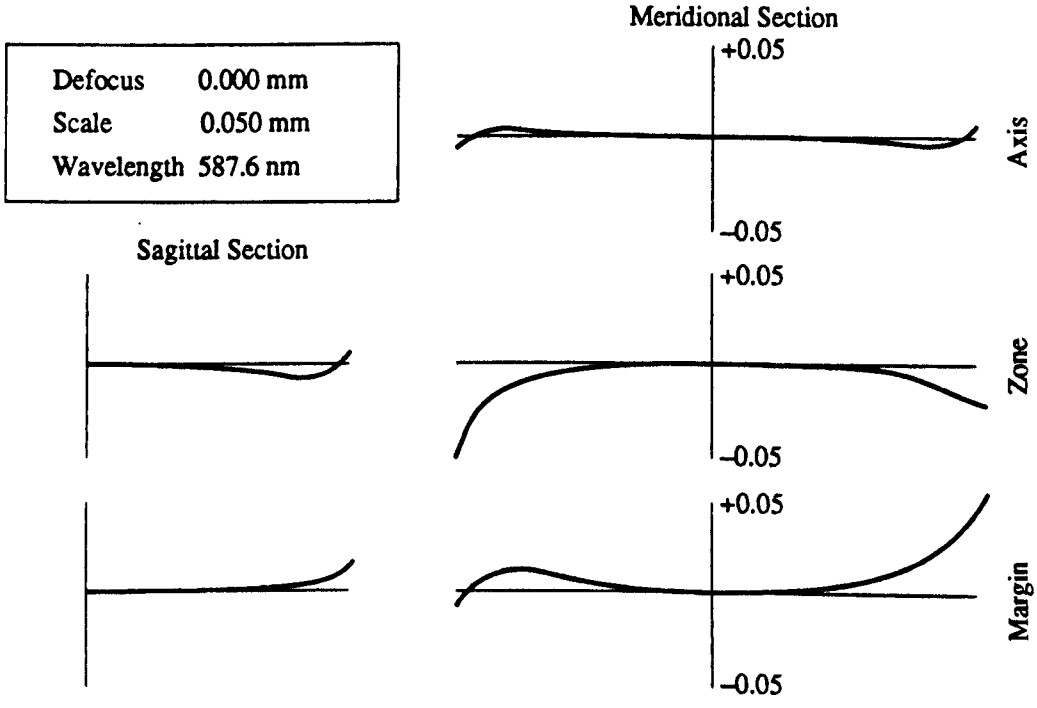


Figure 3.6. Unvignetted transverse ray aberrations for the f/4 50mm triplet of example 3.

of 50mm for example 4 (Appendix C, TL3.2). Although this is a well understood design, it does have a large number of degrees of freedom and there are several possible solutions. Photographic lenses of the double Gauss type should have a flat-field and the distortion should also be small. The approach taken for this correction is to control aberrations sequentially. Although this method can take a larger number of iterations for correction than if all aberrations are corrected simultaneously, there is the advantage that the aberrations which are intractable are found.

The correction proceeds in four stages. Firstly the sagittal focal surface is brought under control while correcting for axial spherical aberration. Secondly, the tangential focal surface is controlled while maintaining the axial correction. The third stage is to target the zonal aperture rays at both off-axis field positions while maintaining the previous correction. The aperture margin rays are left and may assume large values without interfering with the correction. Finally an attempt is made to

reduce the sagittal astigmatism values, although this quickly stagnates with vertex solutions for almost all aberrations. It is worth noting that a damped-least squares run also finds a problem in balancing the s and t focal surfaces to reduce Petzval sum. The resulting aberration curves are shown in in figure 3.7. This design, though, cannot be considered acceptable because of the positive Petzval sum.

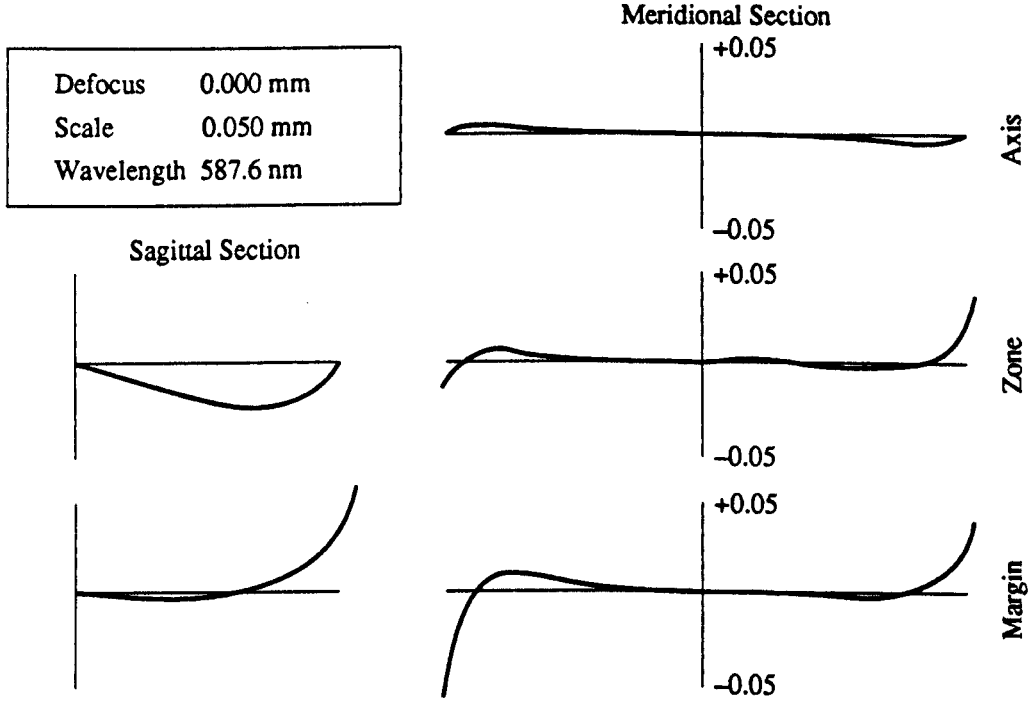


Figure 3.7. Unvignetted transverse ray aberrations for the f/1.7 50mm double Gauss Derivative of example 4.

3.5 Discussion

3.5.1 Merit Function Variation During Correction.

Figures 3.8 and 3.9 show the variation of a sum of squares merit function during the correction of examples 2 and 3 described in the previous section. The merit

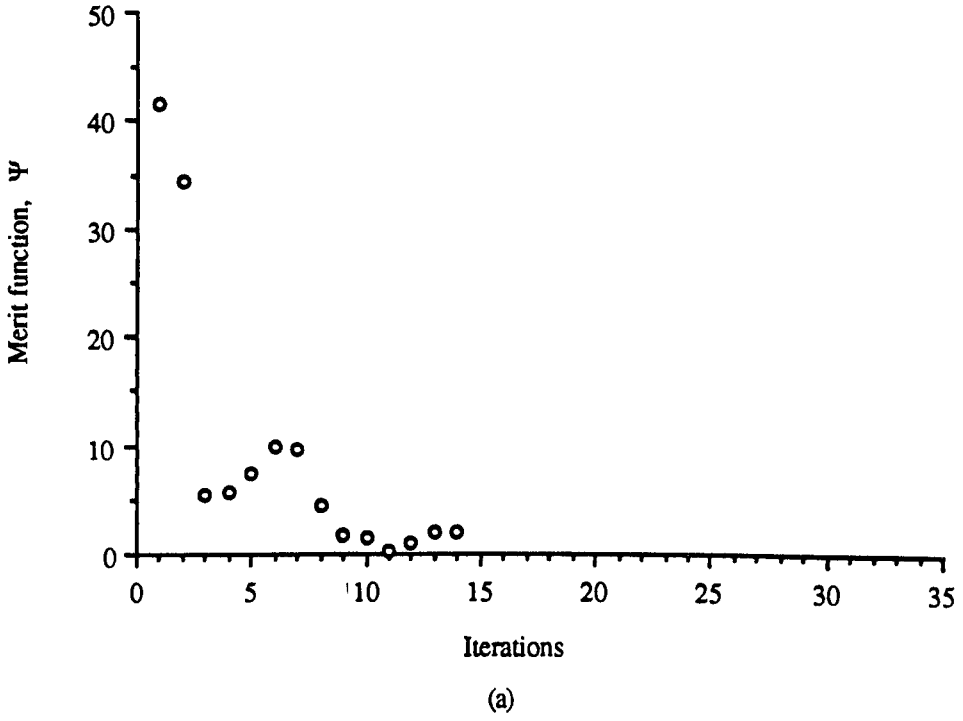
function, Ψ, is defined by

$$\Psi = \sum_{\text{all aberrations}} f_i^2 \tag{3.19}$$

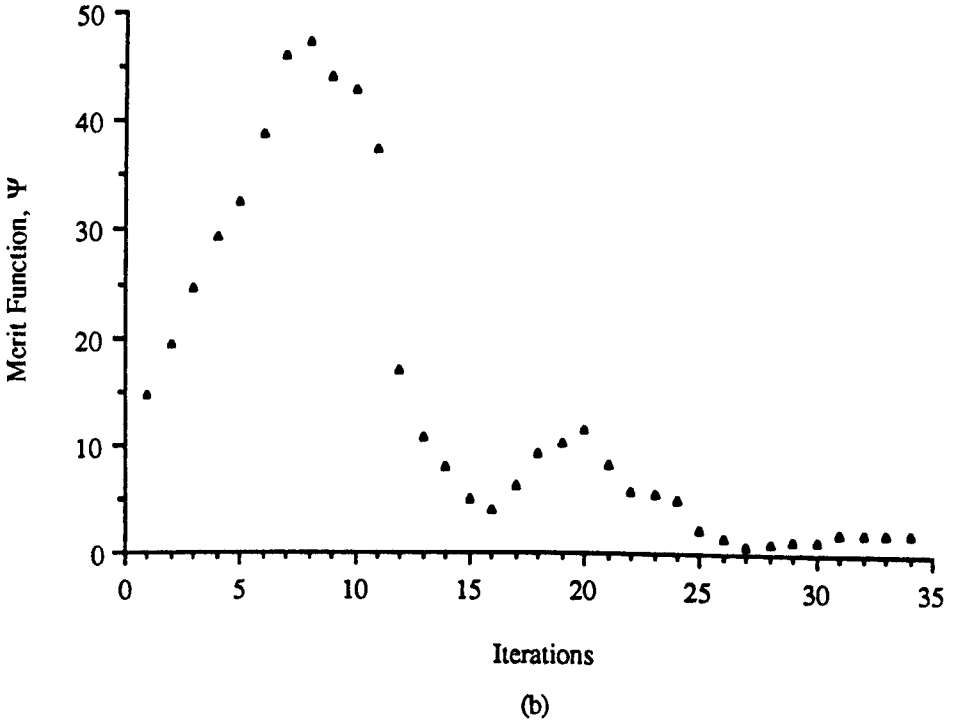
and is therefore the sum of the unweighted squares of the transverse ray aberrations. Any weighting factors would alter the merit function and hence the topography of the merit function surface.

Figures 3.8(a) and 3.9(a) show the merit function variation during quadratic correction of examples 2 and 3 described in §3.4. The general trend in the merit function is downwards, which would be expected since the transverse ray errors usually decrease as the system moves closer to a solution. However, during flat-field correction (fig. 3.8(a)), there are small but clearly evident increases in Ψ. Two possible reasons for this are; firstly, that by targeting only a few specific aberrations, untargeted aberrations can assume large values which dominate the total value of the merit function; and secondly, some aberrations are allowed to drift within a specified tolerance band and therefore may contribute to an increase in Ψ if they assume values towards the upper limit of the band. The increases in Ψ are important since they demonstrate the ability of programs, which do not define or use a merit function, to avoid what would be local minima in least squares based optimization. Figures 3.8(b) and 3.9(b) show the variation of Ψ for identical runs with the linear adaptive method. Although the variation shows identical characteristics to figures 3.8(a) and 3.9(a), the increase in step size and efficiency of the QAC algorithm are clearly evident. These topics are investigated further in §3.5.2 and §3.5.3.

There is another important observation that can be made if the hypothetical merit function is considered further: The QAC algorithm (and the adaptive method), have the ability to alter the aberrations participating in the correction at each iteration. This lends a dynamic nature to the merit function and hence the merit function surface is continually changing its contours. This effect would be hard to quantify, although it is clearly an advantageous feature. It would also appear sensible, since although the target values have been specified by the designer, nothing is known about the direction taken or the topography of the merit function surface.

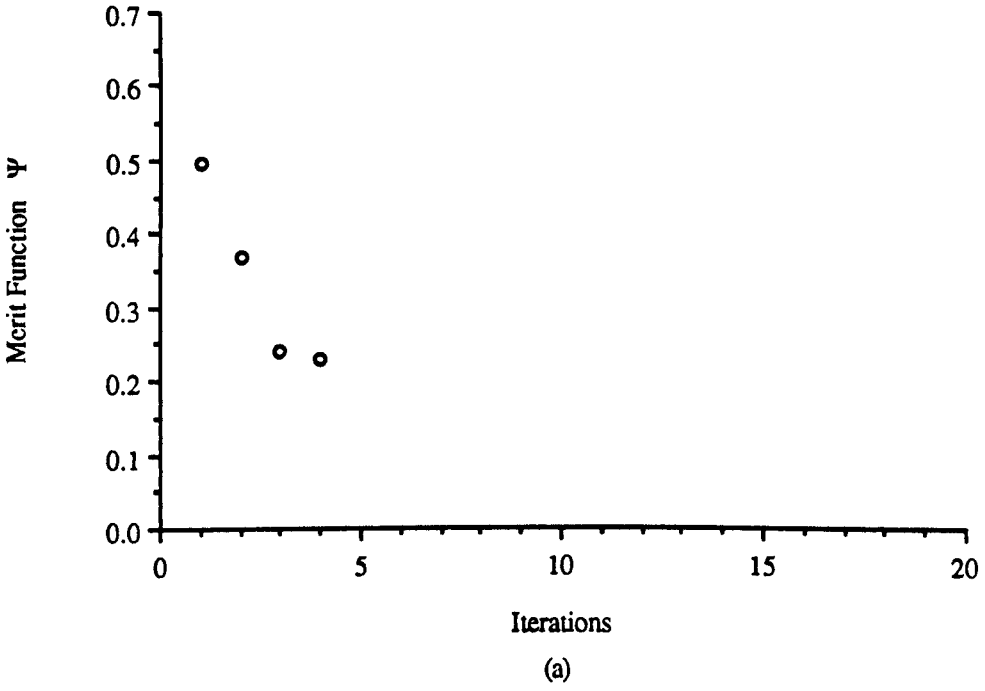


(a)

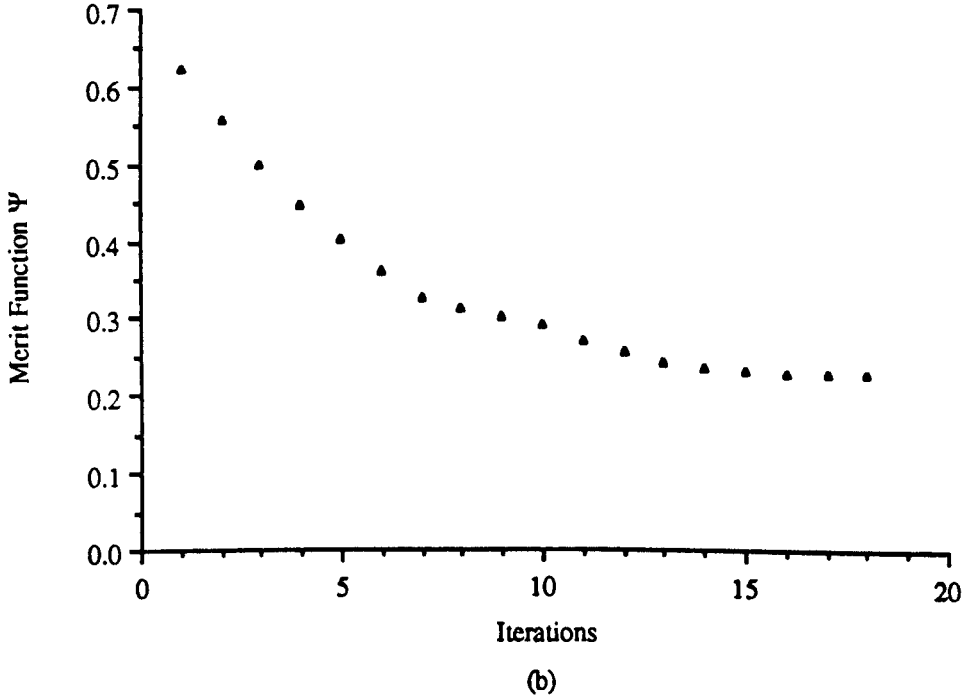


(b)

Figure 3.8 (a) Merit function variation during flat-field correction of an f/4 triplet with the QAC algorithm.
(b) Merit function variation during flat-field correction of an f/4 triplet with the adaptive method.



(a)



(b)

Figure 3.9 (a) Merit function variation during final correction of an f/4 triplet with the QAC algorithm.
(b) Merit function variation during final correction of an f/4 triplet with the adaptive method.

Although the above discussion is instructive, it is somewhat academic since equation (3.7), which is solved to find an initial direction, $\mathbf{x}$, is fundamentally different from the damped least squares solution which is given by

$$\mathbf{x} = -(p^2\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{f_o}. \tag{3.18}$$

where p is the damping factor and $\mathbf{I}$ is the identity matrix. Unless the directions are identical, the comparison of how the merit function variations differ becomes invalid. For this reason, a comparison hasn't been made with a similar DLS optimization.

3.5.2 Relative Computational Efficiency.

In this section the computational efficiency of the quadratic correction algorithm is compared with a version of the linear adaptive method. This has been programmed following the description given by Glatzel and Wilson [3.8]. Although comparisons could be made with other linear algorithms, the adaptive method has the advantage of solving the same minimal equations (equation (3.7)) and is arguably one of the most efficient linear methods in its use of the change matrix. The chosen measure of computing load is the number of complete system raytraces required for a specified correction. A complete system raytrace is defined as the finite raytracing required to assess fully a new or perturbed system; this involves the tracing of seventeen rays: 3 for the axial field position and 7 at each off-axis field position.

Tests have been run in which a comparison of the computational efficiency has been made between the linear adaptive method and the QAC algorithm for the first three of the examples described in §3.4. The results showing the number of complete system raytraces taken by each algorithm are presented in Table 3.5. In addition, an efficiency factor (EF) has been defined as the ratio of the number of system raytraces taken by the adaptive method (AM) to the number taken by the QAC algorithm. Although the results are favourable towards QAC, a large number of intermediate target levels when it is not appropriate, soon negates the increased efficiency. This is because the change matrix is computationally expensive to calculate; whereas an

inner cycle requires a single system assessment, the change matrix requires n system raytraces, where n is the number of variables. As a result QAC usually benefits from fewer intermediate target levels than the adaptive method since the quadratic approximation allows larger steps to be taken. This is investigated further in the next section.

Example	QAC	AM	EF
1. Primary correction of an f/4 triplet	128	427	3.34
2. Flat-field correction of an f/4 triplet	104	462	4.44
3. Final correction of an f/4 triplet	32	46	1.44

Table 3.5. Comparison of computational efficiency between the linear adaptive method and the QAC algorithm.

3.5.3 Average Step Size.

Table 3.6 compares the average length of the applied parameter change vector for the four examples described in §3.4 where the step size is calculated as the vector-2 norm or the Euclidean length of the parameter change vector $\| \mathbf{x} \|$, and these values are averaged over the number of iterations taken for the specified correction.

Example	QAC	AM	Ratio
1. Primary correction of an f/4 triplet	0.3513	0.0361	9.73
2. Flat-field correction of an f/4 triplet	0.0099	0.0017	5.82
3. Final correction of an f/4 triplet	0.0027	0.0005	5.40

Table 3.6. Comparison of the mean step size between the linear adaptive method and the QAC algorithm.

The results show an average increase in step size of magnitude five for the QAC algorithm when compared to the linear adaptive method. However, it is important

to note that, when considering the average applied step length, the results of the previous section are also considered. For example, it would be possible to apply large changes that continually overshoot the solution and at worst could even diverge from it. The fact that *both the step size and computational efficiency* have been increased by similar orders of magnitude, allows us to conclude that the QAC algorithm can successfully make large changes in the system variables.

3.5.4 Comparison with the Adaptive Method.

One of the interesting features of the adaptive method is its efficient use of the change matrix by what Glatzel and Wilson [3.8] refer to as "inner cycles." This is particularly important since the main computational expense is in the calculation of the change matrix, **A**, and the matrix inversion when solving equation (7). The process of inner cycles works by repeatedly using the same change matrix until the linear approximation has been finally exhausted. The linear approximation is therefore only a tangent to the actual variation during the first inner cycle. This has been described in detail by Glatzel and Wilson [3.8].

One advantage of the quadratic approximation method is that it could potentially eliminate the inner cycles of the adaptive method which each require an extra system evaluation. To test this, the correction of the examples described in §3.4 have also been attempted with a version of the adaptive method and table 3.7 shows the number of iterations and inner cycles taken by both the QAC algorithm and the adaptive method. The results demonstrate a marked improvement for the QAC algorithm even though the adaptive method appeared to use the 'inner cycles' successfully. Also the number of iterations taken by the QAC algorithm is significantly less than the adaptive method and therefore not only have the inner cycles been eliminated but an increase in efficiency has also been achieved.

Example	QAC	Adaptive Method	
	itns	itns	inner cycles
1. Primary correction of an f/4 triplet	18	35	137
2. Flat-field correction of an f/4 triplet	15	41	102
3. Final correction of an f/4 triplet	4	3	18

Table 3.7. Comparison of the Number of Iterations Between Adaptive Correction and the QAC Algorithm.

There are several other differences which can be summarised as follows:

The adaptive method described by Glatzel and Wilson [3.9] applies a fraction of the maximum requested single parameter change and continues to do so until another requested parameter change becomes larger, at which point the correction switches to this new parameter. This process continues until all requested parameter changes are zero, at which time the current targets have been reached. Experience with the QAC algorithm has shown that this does not work well since equation (3.7) determines a set of parameter changes whose *combined effect* would be to reach the aberration targets. The lens variables are well known for being non-orthogonal with respect to the aberrations and so only changing one parameter would not necessarily have a positive effect. Rayces [3.15] has also reported similar findings and described a version of the adaptive method where all changes are applied simultaneously.

Finally, the tolerance values for the aberrations are given as percentages of the current target values and this has the effect of narrowing the tolerance band as the targets are reduced to their specified levels. Of course this is not the case when it is desired to increase the target value of the aberration; not a common situation.

3.6 References

[3.1] Gradsteyn and Ryzhik, *Tables of Integrals, Series and Products*, 0.319, pp12, Academic Press (1986).

[3.2] M'Aulay, A.L. and Cruickshank, F.D., "A Differential Method of Adjusting the Aberrations of a Lens System," Proc. Phys. Soc., **57**, pp 302–310 (1945).

[3.3] Cruickshank, F.D., "A System of Transfer Coefficients for Use in the Design of Lens Systems," Proc. Phys. Soc., **57**, pp 350–361 (1945).

[3.4] MacDonald, J. and Yang, Y., "Analytical Derivatives in Damped-Least Squares Optimization," in Proc. SPIE Vol. **655**, *Optical System Design, Analysis and Production fior Advanced Technology Systems* , pp 49–53 (1986).

[3.5] Huber, E.D., "Extrapolated least squares optimization in optical design," JOSA A, **2**, pp 544–554 (1985).

[3.6] see for example, descriptions of the Fletcher-Reeves conjugate gradient method or the Davidon-Fletcher-Powell variable metric (sometimes referred to as quasi-Newton) method which are desribed in a number of texts on numerical analysis and optimization. Both these techniques are iterative and build up information about the Hessian matrix from preceding iterations since it cannot be evaluated explicitly.

[3.7] Maxwell, J. and Hull, C.C., "A Simple Multidimensional Quadratic Extrapolation Method for the Specific Correction of Aberrations in Optical Systems," in Proc. SPIE Vol. **1354**, *International Lens Design Conference* , pp 277–285 (1990).

[3.8] Glatzel, E. and Wilson, R. "Adaptive Automatic Correction in Optical Design," Appl. Opt., **7**, 265–276 (1968)

[3.9] Peressini *et al.*, *The Mathematics of Non-Linear Programming*, Springer-Verlag, New York (1980).

[3.10] Press, W.H. *et al.*, *Numerical Recipes*, Cambridge University Press, Cambridge (1986), Chap. 2.

[3.11] Atkinson, K.E., *Introduction to Numerical Analysis*, John Wiley, New York (1978) pp 450–454.

[3.12] Reference [3.10], pp 52–64.

[3.13] Feder, D.P., "Automatic Lens Design Methods", J.O.S.A., **47**, 902–912 (1957).

[3.14] Tatian, B., "Nonlinearity and Lens Design", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, pp154–164 (1990)

[3.15] Rayces, J.L. "Ten Years of Design with Glatzel's Adaptive Method," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, pp75–84 (1980)

CHAPTER 4

Generating Starting Designs with Higher Order Aberration Correction

4.1 Introduction

The first stage in the design process, after selecting an initial lens configuration, is usually to reduce the larger primary aberrations. In particular, it is generally understood that the reduction of the large edge-of-aperture or edge-of-field aberrations is an important step before assessing the design's potential for well balanced aberrational characteristics. However, optical design optimization programs are ill-suited to carrying out this initial correction stage since they perform most efficiently on well corrected starting designs. The principal reason for this is that if large parameter changes have to be made, it is possible that multiple local minima solutions will be encountered which will prevent the algorithm reaching a more desirable region of parameter space.

The higher order aberrations generally have slower rates of change than their third order counterparts and therefore large parameter changes are required to correct them. This is in direct opposition with the requirement to reduce the larger (primary) aberrations since these require comparatively small parameter changes. Such changes would tend not to affect the higher order contributions significantly.

In this chapter, the initial correction of an optical system is attempted by correcting the higher order contributions prior to reducing the primary aberrations. The

algorithm developed in Chapter 3 is used for this preliminary correction since it has been shown to be capable of making large parameter changes without becoming trapped in local minima solutions. Although this approach to preliminary correction has, to the author's knowledge, not been reported in the literature, Rogers [4.1] has indicated that similar approaches have been investigated by other workers in the field.

4.2 Mixed Higher Order Aberrations

4.2.1 Mixed Higher Order Monochromatic Aberrations

Mixed higher order aberrations, which contain all contributions to a particular aberration of fifth order and above, have been used for a number of years as an approximation for evaluating the magnitude of the higher order contributions (see, for example, Cox [4.2]). However, little practical use has been made of them during the correction of optical systems. They can be calculated from finite ray trace data and a knowledge of the third order aberrations. Expressions for the mixed higher order transverse ray aberrations, which are required later in the chapter, are developed in this section.

The finite transverse ray aberrations can be subdivided into three constituent components: aberrations which are even functions of the aperture coordinate, aberrations which are odd functions with respect to aperture and finally those aberrations which result in the transverse ray aberration curves having a finite gradient at their origin (the point at which the finite principal ray strikes the image plane). These groupings correspond to the coma terms, the spherical aberration-like terms and the astigmatism and field curvature terms of the total ray aberration. Distortion is an odd function of the aperture coordinate, but it is a special case because it is not usually represented as a transverse ray aberration. This is because the *finite* principal ray is usually chosen as the reference for calculating the transverse ray

aberrations; any difference between the paraxial and finite principal ray heights at the image plane would just add a constant term to all the aberration values for that particular field position.

In the following analysis, the finite transverse ray aberrations in the meridian and sagittal sections are denoted by $\delta\eta(r,\varrho)$ and $\delta\xi(r,\varrho)$ respectively, where r is the normalized aperture coordinate and ϱ the relative field. $\delta\eta(r,\varrho)$ in the sagittal section is ignored here as is commonly done on transverse ray aberration curve plots. For the axial field position, only spherical aberration terms are present and the defocus is assumed to be zero. Therefore, the finite ray aberration is the sum of the spherical aberration terms alone and hence the finite ray aberration for the axial marginal ray can be written as

$$\delta\eta(1,0) = \sum_{\text{all orders}} \text{Spherical Aberration Terms} \tag{4.1}$$

However, the transverse component of third order spherical aberration is given by $S_I/2nu'$ (see, for example, [4.3]) where S_I is the Seidel coefficient for spherical aberration which can be computed from a paraxial raytrace of the system and u' is the convergence angle of the paraxial axial marginal ray. If a mixed higher order aberration coefficient, M_I, is defined as the sum of all the spherical aberration terms of fifth order and above then from (4.1)

$$M_I = \delta\eta(1,0) - \frac{S_I}{2nu'}. \tag{4.2}$$

This aberration is field and aperture azimuth invariant and can be subtracted out from subsequent calculations at off-axis field positions.

The full semi-field position, $\varrho = 1$, is chosen for calculation of the other mixed higher order aberrations, since the Seidel aberrations are usually scaled to this field position, thereby facilitating the subtraction of the third order contributions. In the sagittal section, only aberrations which are odd functions of the aperture coordinate are present. These include the field invariant spherical aberration which has already been evaluated, oblique spherical aberration (OBSA) and sagittal astigmatism and

field curvature (FC). Therefore the finite transverse ray aberration at the edge of the aperture in the sagittal section can be written as

$$\delta\xi(1,1) = \sum_{\text{all orders}} \text{SA} + \text{OBSA} + \sum_{\text{all orders}} \text{Astigmatism and FC} . \qquad (4.3)$$

The astigmatism and field curvature term may be found by tracing s & t foci along the principal ray using Young's formulae [4.4] (sometimes referred to as the Coddington equations). The result of applying Young's equations are finite aberration values for the sagittal and meridional astigmatism, which are denoted by s' and t' respectively. Equation (4.3) may now be rewritten as

$$\delta\xi(1,1) = \delta\eta(1,0) + \text{OBSA} + s' \qquad (4.4)$$

and defining M_{VI} as the mixed higher order aberration relating to oblique spherical aberration,

$$M_{VI} = \delta\xi(1,1) - \delta\eta(1,0) - s'. \qquad (4.5)$$

The astigmatism and field curvature term may also be subdivided into its third order and higher order components by subtracting the primary aberration contribution from the results of the s & t trace. Therefore a mixed higher order aberration M_{IV} can be defined by

$$M_{IV} = s' - \frac{S_{III} + S_{IV}}{2nu'} \qquad (4.6)$$

where $(S_{III}+S_{IV})/2nu'$ is the transverse primary aberration component of sagittal astigmatism and field curvature.

The meridian section contains aberrations which are both odd and even functions of the aperture coordinate, r. If these are denoted by $\delta\eta_o$ and $\delta\eta_e$ respectively then from figure 4.1

$$\delta\eta(1,1) = \delta\eta_e + \delta\eta_o \qquad (4.7)$$

$$\delta\eta(-1,1) = \delta\eta_e - \delta\eta_o \qquad (4.8)$$

which can be rearranged to give

$$\delta\eta_e = \tfrac{1}{2}\{\delta\eta(1,1) + \delta\eta(-1,1)\} \qquad (4.9)$$

$$\delta\eta_o = \tfrac{1}{2}\{\delta\eta(1,1) - \delta\eta(-1,1)\}. \qquad (4.10)$$

The even component of the aberration contains the coma components of the finite transverse ray aberration,

$$\delta\eta_e = \sum_{\text{all orders}} \text{coma} \tag{4.11}$$

and therefore subtracting the primary aberration component will leave the mixed higher order term, M_{II}, which is given by

$$\begin{aligned} M_{II} &= \delta\eta_e - \frac{3S_{II}}{2nu'} \\ &= \tfrac{1}{2}\{\delta\eta(1,1) + \delta\eta(-1,1)\} - \tfrac{3S_{II}}{2nu'}. \end{aligned} \tag{4.12}$$

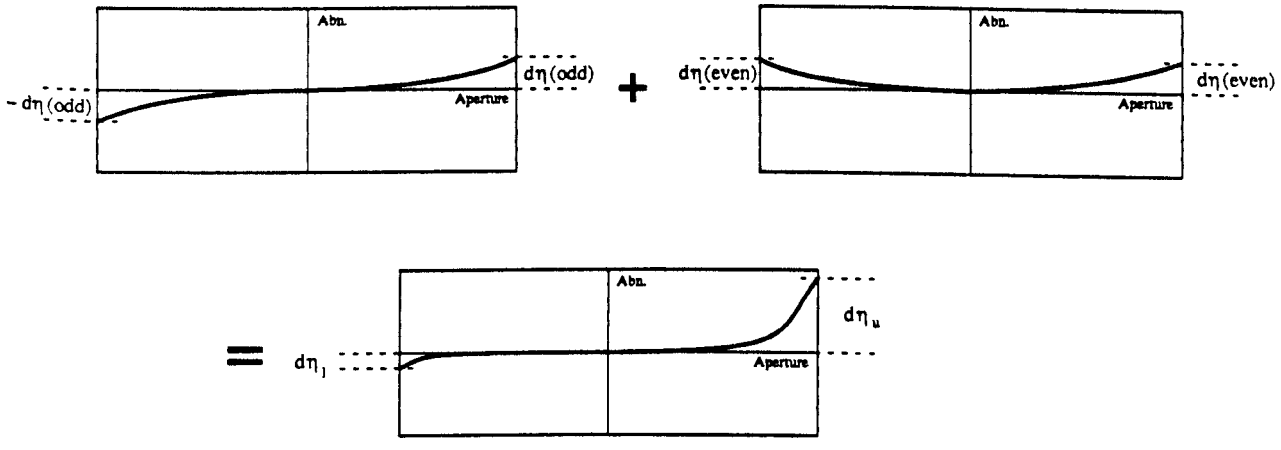


Figure 4.1 Even and odd function contributions to the finite transverse ray aberration.

The odd component, $\delta\eta_o$ of the finite ray aberration consists of spherical aberration, a higher order astigmatism term (HOA) and tangential astigmatism and field curvature components. Therefore,

$$\delta\eta_o = \sum_{\text{all orders}} \text{SA} + \text{HOA} + \sum_{\text{all orders}} \text{Astigmatism and FC.} \tag{4.13}$$

Following an identical approach to that used for the sagittal section, the finite astigmatism and field curvature term can be evaluated by tracing the t focus along

the finite principal ray and hence substituting from equations (4.10) and (4.1), equation (4.13) becomes

$$\tfrac{1}{2}\{\delta\eta(1,1) - \delta\eta(-1,1)\} = \delta\eta(1,0) + \text{HOA} + t' \tag{4.14}$$

where t' is the finite transverse ray aberration component of the tangential astigmatism and field curvature. Defining M_{VII} as the mixed higher order aberration due to higher order astigmatism, then

$$M_{VII} = \tfrac{1}{2}\{\delta\eta(1,1) - \delta\eta(-1,1)\} - \delta\eta(1,0) - t'. \tag{4.15}$$

The astigmatism and field curvature term may be subdivided by subtracting the primary aberration component $(3S_{III} + S_{IV})/2nu'$ to give the mixed higher order term,

$$M_{III} = t' - \frac{(3S_{III} + S_{IV})}{2nu'}. \tag{4.16}$$

Distortion is not usually represented as a transverse ray aberration since it would simply add a constant term to all aberration values for a given field position. The finite ray aberration due to distortion is measured as the difference between the heights at which the paraxial principal ray $\bar{y}'$ and the finite principal ray $\bar{Y}'$ strike the image plane. Therefore, the finite transverse error due to distortion $\delta\bar{\eta}$ is given by

$$\delta\bar{\eta} = \bar{y}' - \bar{Y}' \tag{4.17}$$

and hence the mixed higher order component of distortion, M_V is given by

$$M_V = \delta\bar{\eta} - \frac{S_V}{2nu'} \tag{4.18}$$

where $S_V/2nu'$ is the primary aberration component of transverse distortion. The seven mixed higher order aberration coefficients that have been defined are given in table 4.1.

Aberration	Value
M_I	$\delta\eta(1,0) - S_I/2nu'$
M_{II}	$\frac{1}{2}\{\delta\eta(1,1) + \delta\eta(-1,1)\} - 3S_{II}/2nu'$
M_{III}	$t' - (3S_{III} + S_{IV})/2nu'$
M_{IV}	$s' - (S_{III} + S_{IV})/2nu'$
M_V	$\delta\bar{\eta} - S_V/2nu'$
M_{VI}	$\delta\xi(1,1) - \delta\eta(1,0) - s'$
M_{VII}	$\frac{1}{2}\{\delta\eta(1,1) - \delta\eta(-1,1)\} - \delta\eta(1,0) - t'$

Table 4.1 Mixed Higher order Transverse Ray Aberrations.

4.2.2 Extension to Chromatic Aberrations

The analysis performed in the previous section, where the third order contributions were subtracted from the finite ray aberrations to find the mixed higher order contributions, becomes more complicated when applied to chromatic aberrations. This is because the two primary chromatic aberrations - chromatic variation of back focal length and magnification respectively - have no higher order counterparts. For example, the ray aberration associated with chromatic variation of focal length would, following the approach of the previous section, be evaluated by tracing a finite axial marginal ray at the blue and the red wavelengths. However, this ray not only contains information about chromatic variation of back focal length but also about axial spherochromatism. It would be possible to evaluate the spherochromatism by computing the Seidel aberration coefficient, S_I, at the blue and red wavelengths and the difference between these two values would give a measure of the chromatic variation of primary spherical aberration.

The subject of chromatic correction and the associated area of glass selection are outside the scope of this thesis and therefore the development of mixed higher order chromatic aberrations has not been attempted here.

4.3 Reduction of the Higher Order Aberrations

4.3.1 Simultaneous Correction of Higher Order Aberrations

Simultaneous correction of all the higher order aberrations is the first stage of the initial correction process as described in this chapter. The success of this method is dependent on the relative rates of change of the chosen aberrations; the aberrations with slow rates of change will tend to have the greatest influence on the direction of the solution vector and hence the aberrations with faster rates of change will have a tendency to overshoot their target values. As a result, the tolerance criteria for each

intermediate aberration target value is modified from that described in Chapter 3 §3.3.

The objective of the initial correction process which is being used here is to reduce the magnitude of the higher order aberrations and therefore at each intermediate target level a test is made to see if the current aberration value is smaller in magnitude than its current target value plus the tolerance, that is;

$$|(f_i)_{current}| < |(f_i)_{target}\{1 + tol/100\}| \tag{4.19}$$

where *tol* is the specified tolerance on the aberration expressed as a percentage. This criterion allows an aberration to overshoot its target value provided the magnitude of the aberration remains low. This technique has been found to be of advantage in any situation requiring the reduction of the magnitude of aberrations.

4.3.2 The Effect of Boundary Conditions

It is possible to reduce all the mixed higher order aberrations simultaneously by targeting each aberration to the required level. At the same time the paraxial specification (EFL and numerical aperture) must be maintained and several methods are available for this: Scaling the lens data after each iteration would appear to be desirable since it removes a paraxial quantity from the change matrix whose presence has been shown to significantly increase the degree of ill-conditioning (§3.3.3). However, in practice it can lead to oscillation of aberrations about their target values and it also tends to destroy any correction that has been achieved (fig. 4.2).

This is further aggravated by the large changes required when modifying the higher order aberrations. It is therefore necessary to target the EFL or the paraxial marginal ray angle u' to maintain the paraxial specification and to tolerate the resulting increase in ill-conditioning. Figure 4.2 shows the variation of M_{III} during the correction of all mixed higher order aberrations using u' to control the EFL (labeled u') and scaling the lens data after each iteration to maintain the paraxial

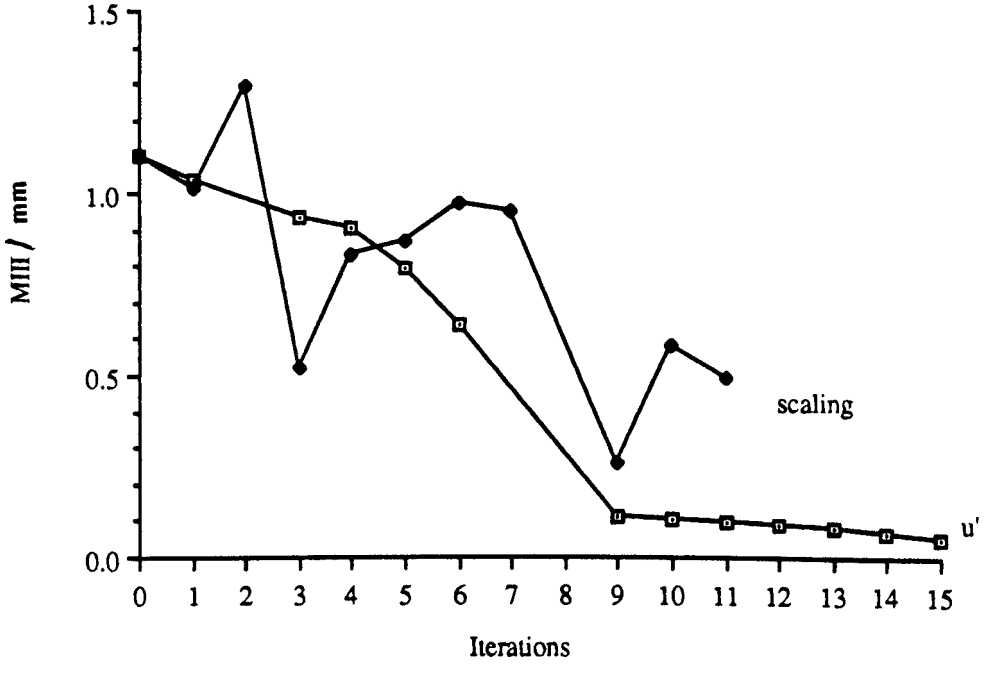


Figure 4.2 Variation of the mixed higher order aberration M_{III} using u' to control EFL and by scaling the lens data.

specification (labeled "scaling"). It can be seen from the figure that if scaling is used to control the paraxial specification, progress towards the target value (0.05) is erratic and that oscillation is also evident. However, using u' the reduction is more controlled with no evidence of oscillation about intermediate target levels.

4.3.3 Grouped Correction of Higher Order Aberrations

In situations where combined correction is unsuccessful, it is often possible to reduce the higher order aberrations by building up the matrix of targeted aberrations. To do this in anything more than an *ad hoc* manner, the rates of change of the aberrations must be examined and aberrations grouped together with similar rates of change. For this purpose, a measure of the average rate of change of an aberration has been defined as

$$RRC = \frac{\sqrt{\sum_{i}^{m} a_{ij}^2}}{m} \tag{4.20}$$

where $a_{ij} = \partial f_i / \partial x_j$ for aberrations $f_i, i = 1 \ldots m$ and variables $x_j, j = 1 \ldots n$. (RRC is an abbreviation for row rate of change).

Grouped correction is performed by targeting the slowly varying aberrations first (usually only two or three, typically M_{IV}, M_V and M_{VI}) and reducing them as far as possible before targeting the remaining aberrations. If there are very disparate rates of change then an additional group may be necessary. In this way, the aberrations that require the largest parameter changes are corrected first, with subsequent groups requiring smaller changes in the system which do not significantly upset any previous correction. This method requires that the ratio between the various rates of change of the aberrations should remain constant as the correction proceeds.

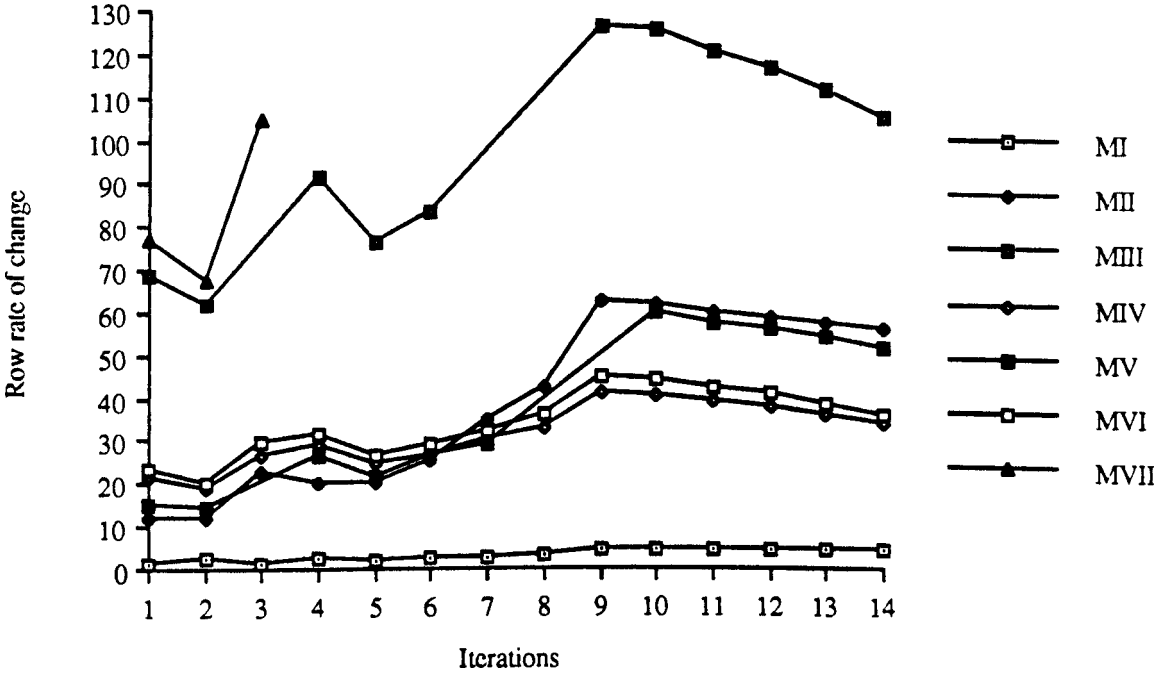


Figure 4.3 Variation of the row rates of change of all seven mixed higher order aberrations during correction of an f/2 Petzval lens.

From figure 4.3, which shows the variation in the row rates of change of the mixed higher order aberrations during correction of an f/2 Petzval lens which is described in §4.4.4, it can be seen that the variation of all seven mixed higher order aberrations follows a common trend and that the ratios between the aberrations remains approximately constant. Therefore the method of grouped higher order aberration correction described above is reasonable, based on the observations shown in figure 4.3.

4.3.4 Variation of the Seidel Aberrations

The program described in §3.3 of chapter 3 has the potential to treat the Seidel aberrations as controlled aberrations, therefore preventing them from assuming very high values. This is achieved by only targeting an aberration if it drifts outside a specified tolerance band. In the examples that have been corrected for their higher order aberrations (§4.4), it has not been found necessary to target or control the Seidel aberrations at all; they can be allowed to assume any value. The variation of the Seidel aberrations for combined correction of all mixed higher order aberrations in an f/2 Petzval lens is shown in figure 4.4.

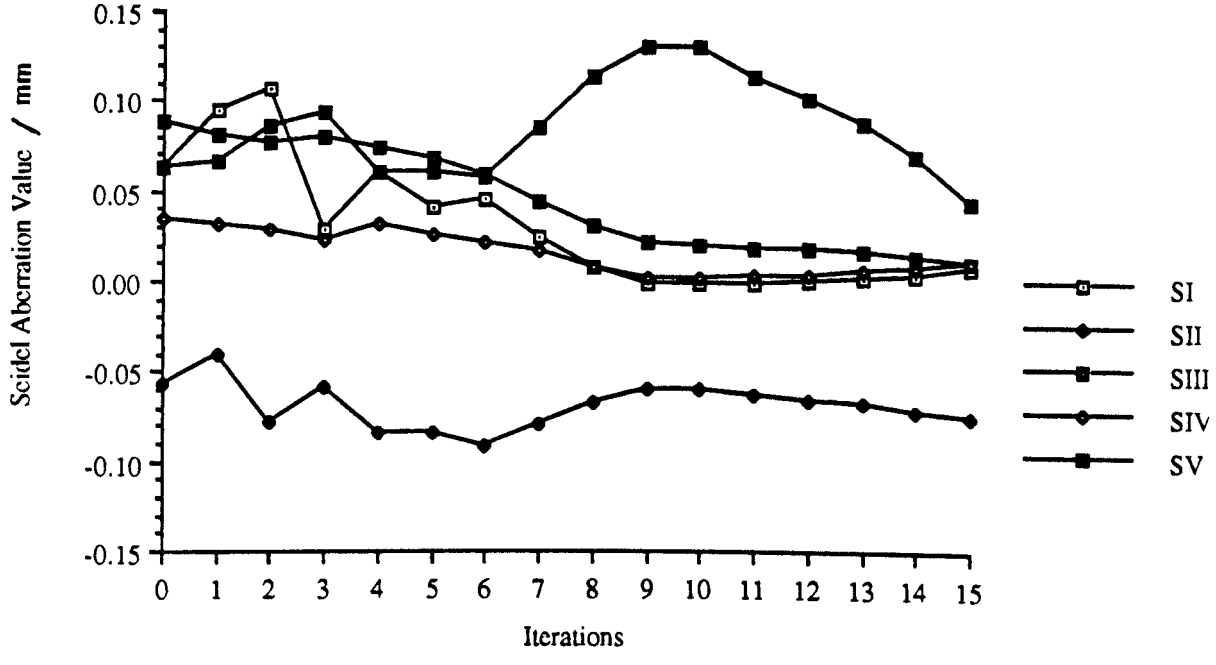


Figure 4.4 Variation of the Seidel aberrations during combined higher order correction of an f/2 Petzval lens.

Figure 4.4 shows that there does not appear to be any tendency for the Seidel aberrations to diverge and therefore it is unlikely that they will attain very high values. This has been confirmed by the examples in §4.4. It is possible that targeting the higher order aberrations to small values ensures that the design is reasonable and therefore no single aberration value becomes excessively large.

4.3.5 Computer Implementation

It is not possible to predict what magnitudes the higher order aberrations can be reduced to and therefore two simple modifications have been made to the program described in chapter 3 that prove essential during this 'blind' correction. First, the program reduces the aberrations to their specified target values via a set of intermediate or reduced aberration target levels whose number is specified by the user. As the aberrations reach these intermediate target values, a file is saved which contains a partially corrected lens. Should no further improvement be possible then this partially corrected design can be recovered for further correction or assessment. Second, the lens is saved at the start of every correction run; it is therefore always possible to recover the lens data if a run diverges or, in the very worst case, causes a program crash.

4.3.6 Choice of Variables

The choice of variables is constrained by the fact that there must be at least as many variables as aberrations during each iteration. The maximum number of aberrations at each iteration is eight; seven mixed higher order aberrations (M_I-M_{VII}) and the paraxial marginal ray angle u' which is used to control the EFL. The use of all available curvatures is almost always useful although separations must also be used when they contribute to the aberration correction. However, separations generally have a much slower rate of change than curvatures and therefore their inclusion may lead to ill-conditioning and poor convergence, particularly if they are not useful variables.

4.4 Examples

In this section comparisons are made between the solutions obtained using damped least squares (DLS) optimization on a starting design both before and after it has undergone higher order aberration correction. Changes in efficiency and the final solution are noted. All the example lenses used here have moderate to high apertures (f/4 – f/1) so that the higher order aberrations are significant. The DLS optimization (Kidger Optics Sigma PC), used the weighted sum of the squares of the transverse aberrations as the merit function with a default set of weighting factors. The merit function also included boundary violations (paraxial constraints, edge violations etc.) The default weights vary with field and so merit function comparisons can only be made for designs with the same field specifications.

The objective of this study was to determine whether preliminary higher order aberration correction allowed different solution forms to be found. The designs resulting from a single run with DLS and using the default weighting factors can not be considered well corrected and so aberration curves have not been given.

4.4.1 Experiments with a 50mm f/4 triplet - starting design # 1

The first starting design (# 1) is a 50mm f/4 Cooke triplet comprising two positive plano-convex lens elements placed symmetrically about a biconcave negative lens. The air spaces are each 5mm, the field ±15 degrees with the stop at the second surface of the negative element. The glasses were LAK9 for the outer positive elements and SF8 for the negative lens (Appendix D, TL4.1). This design is close to the final Cooke triplet form and therefore it is not necessarily expected that different solutions will be found.

DLS optimization of the starting design ($\Psi = 430$) with all six curvatures and the two air spaces variable and using multiplicative damping stopped after 16 iterations with a merit function of $\Psi = 74.9$. This lens form, shown in figure 4.5, consists

of three elements in two groups - an airspaced doublet followed by a plano-convex singlet. DLS optimization of starting design # 1 with only the curvatures variable resulted in a symmetrical triplet form with a merit function of $\Psi = 4.45$ after 20 iterations. It is unlikely that this design is an optimum since it suffers from astigmatism and field curvature and has not achieved longitudinal achromatism.

Combined higher order aberration correction of starting design # 1 was attempted with all six curvatures and the two air spaces as variables. The target values for all the mixed higher order aberrations were 0.05. The correction could not achieve better than 4/11ths of the specified target values; the values of the higher order aberrations both before and after correction are given in table 4.2.

Aberration	Start	pc4-11
M_I	-0.00728	0.01456
M_{II}	0.04298	0.04230
M_{III}	0.79606	0.49152
M_{IV}	0.43693	0.27319
M_V	-0.39987	-0.23617
M_{VI}	-0.45332	-0.26548
M_{VII}	-0.86153	-0.52516

Table 4.2 Mixed Higher order Transverse Ray Aberrations both before ('start') and after ('pc4-11') combined higher order correction of an f/4 triplet (starting design # 1).

Combined higher order aberration correction has made two noticeable changes: the power of the negative lens has been increased and the first air space has been reduced to give an asymmetric design. A DLS run on this lens with all six curvatures and the two air spaces variable stopped after only 9 iterations with a merit function of 1.51. Not only was the final merit function lower but there was also a marked increase in efficiency (fig. 4.6). The resulting design is an asymmetric triplet with the negative lens closer to the first positive lens. It suffers from field curvature, which was found to be a common problem when using the default weight settings.

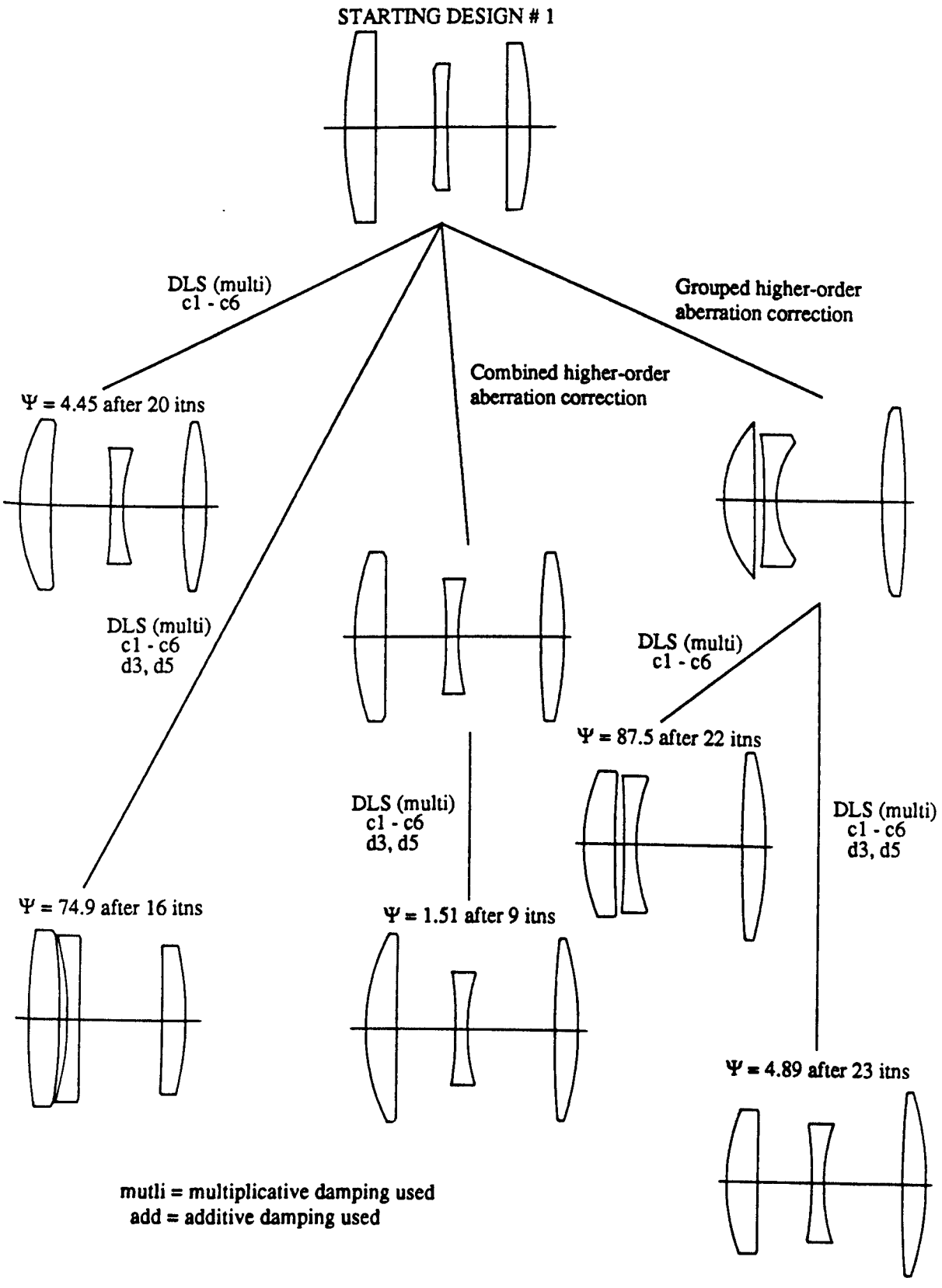


Figure 4.5 Diagrammatic representation of the lenses resulting from the corrections performed on triplet starting design # 1 (see text).

No definite conclusions can be made about the significance of higher order correction from the above example since only partial correction was achieved. It may have just been fortuitous that the superior asymmetric design was found and that there was an increase in efficiency.

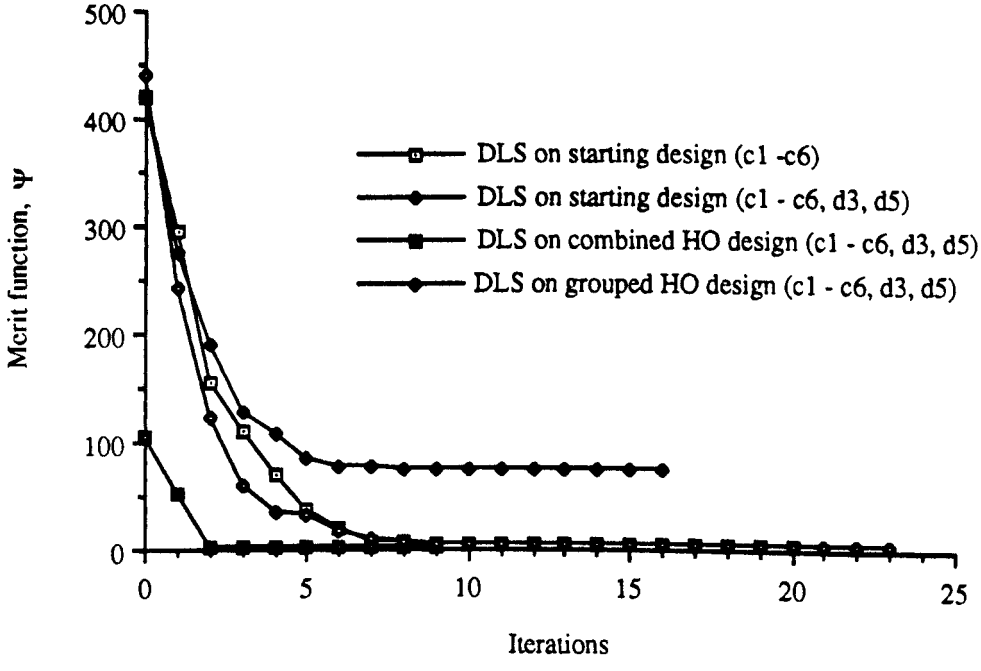


Figure 4.6 DLS optimization runs on starting design # 1 and higher order corrected designs.

To test the significance of higher order aberration correction further on this design grouped higher order aberration correction was attempted (§4.3.2). First the RRC's were examined (table 4.3) and the three slowest aberrations, M_I, M_{IV} and M_{VI}, were targeted. M_{II} was not included because previous experience with this lens in Chapter 3, §3.4 had shown that there was a difficult coma balance with respect to the field. Although the use of *a priori* information makes the process of initial correction far from automatic, the purpose here was to test the effect of significantly reducing the higher order contributions before attempting optimization with DLS.

Aberration	Row rate of change
M_I	0.19
M_{II}	1.05
M_{III}	5.64
M_{IV}	2.11
M_V	4.42
M_{VI}	2.26
M_{VII}	5.97

Table 4.3 RRC's for starting design # 1.

After successful correction of the first group of aberrations to magnitudes of less than 0.05, a second group containing all mixed higher order aberrations with the exception of M_{II} was also corrected to this level. Finally M_{II} was included and all mixed higher order aberrations attained their target value of 0.05 to within the specified tolerance of 50%. Table 4.4 presents the values of the mixed higher order aberrations both before and after grouped higher order aberration correction.

Aberration	Start	Corrected
M_I	-0.00728	0.05640
M_{II}	0.04298	0.04650
M_{III}	0.79606	0.04546
M_{IV}	0.43693	0.04857
M_V	-0.39987	0.05023
M_{VI}	-0.45332	0.05224
M_{VII}	-0.86153	0.05365

Table 4.4 Mixed Higher order Transverse Ray Aberrations both before ('start') and after ('corrected') grouped higher order correction of an f/4 triplet (starting design # 1).

The higher order correction has reduced the first air space to produce a lens with three elements in two groups (fig. 4.5) of similar form to the lens resulting from the initial DLS run on the starting design (# 1) with all curvatures and the two air spaces variable. However, DLS optimization of the design resulting from grouped

higher order aberration correction with all curvatures and the two air spaces variable gave an asymmetric triplet with a merit function of $\Psi = 4.89$ after 23 iterations. The slightly higher merit function is due to an inferior correction of transverse chromatic aberration. This optimization run is plotted in figure 4.6. A second DLS run was made on the same design but with only the curvatures as variables. This resulted in a lens with three elements in two groups and with $\Psi = 87.5$. However, the airspaced achromatic doublet which forms the first group has a nearly plane parallel air space (fig. 4.5). The design of similar form which resulted from DLS optimization of the starting design (# 1) with all curvatures and the two air spaces variable has an air lens between the first two elements which helps to achieve correction of longitudinal chromatic aberration and spherochromatism.

4.4.2 Experiments with a 50mm f/4 triplet - starting design # 2

The starting design used in the previous section was reasonably close to the Cooke triplet form and its correction does not present a very great challenge for many modern optimization programs although several different 'solutions' were found. Starting design # 2, which is used in this section, is a triplet with the same specifications and glasses as starting design # 1 but it consists of three equispaced plane parallel plates with an angle solve on the final surface. The air spaces were each set to 5mm and the stop was again at surface 4 (Appendix D, TL4.2).

DLS optimization of this design with all curvatures and the two air spaces as variables reached an asymmetric solution (negative element closer to the first positive lens) with $\Psi = 1.1$ after 47 iterations (fig. 4.7). This was perhaps surprising since starting design # 1, which is potentially much closer to the Cooke triplet form, converged to a design with three elements in two groups shown in figure 4.5. DLS optimization with just the curvatures variable produced a symmetrical design with $\Psi = 2.58$ after 22 iterations. Both these designs are shown on figure 4.8 which presents all the results for the correction of starting design # 2 and its derivatives.

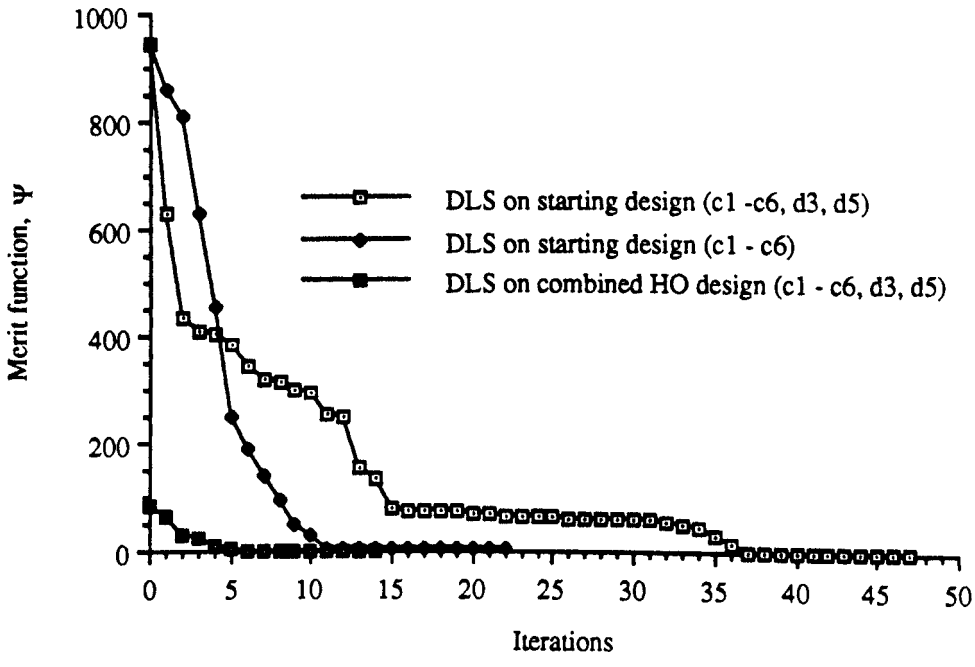


Figure 4.7 DLS optimization runs on starting design # 2.
and higher order corrected designs.

Combined correction of all seven mixed higher order aberrations achieved 7/10ths of the requested aberration target values (0.05). The values of the aberrations both before and after correction are given in table 4.5 although some aberrations have been removed from the matrix due to intractability and ill-conditioning.

Aberration	Start	Corrected
M_I	-0.01603	0.00469
M_{II}	0.09225	0.05494
M_{III}	0.77339	0.28408
M_{IV}	0.42393	0.16176
M_V	0.03961	-0.46559
M_{VI}	-0.44349	-0.10059
M_{VII}	-0.34954	-0.14746

Table 4.5 Mixed Higher order Transverse Ray Aberrations both before ('start') and after ('corrected') combined higher order aberration correction of an f/4 triplet (starting design # 2).

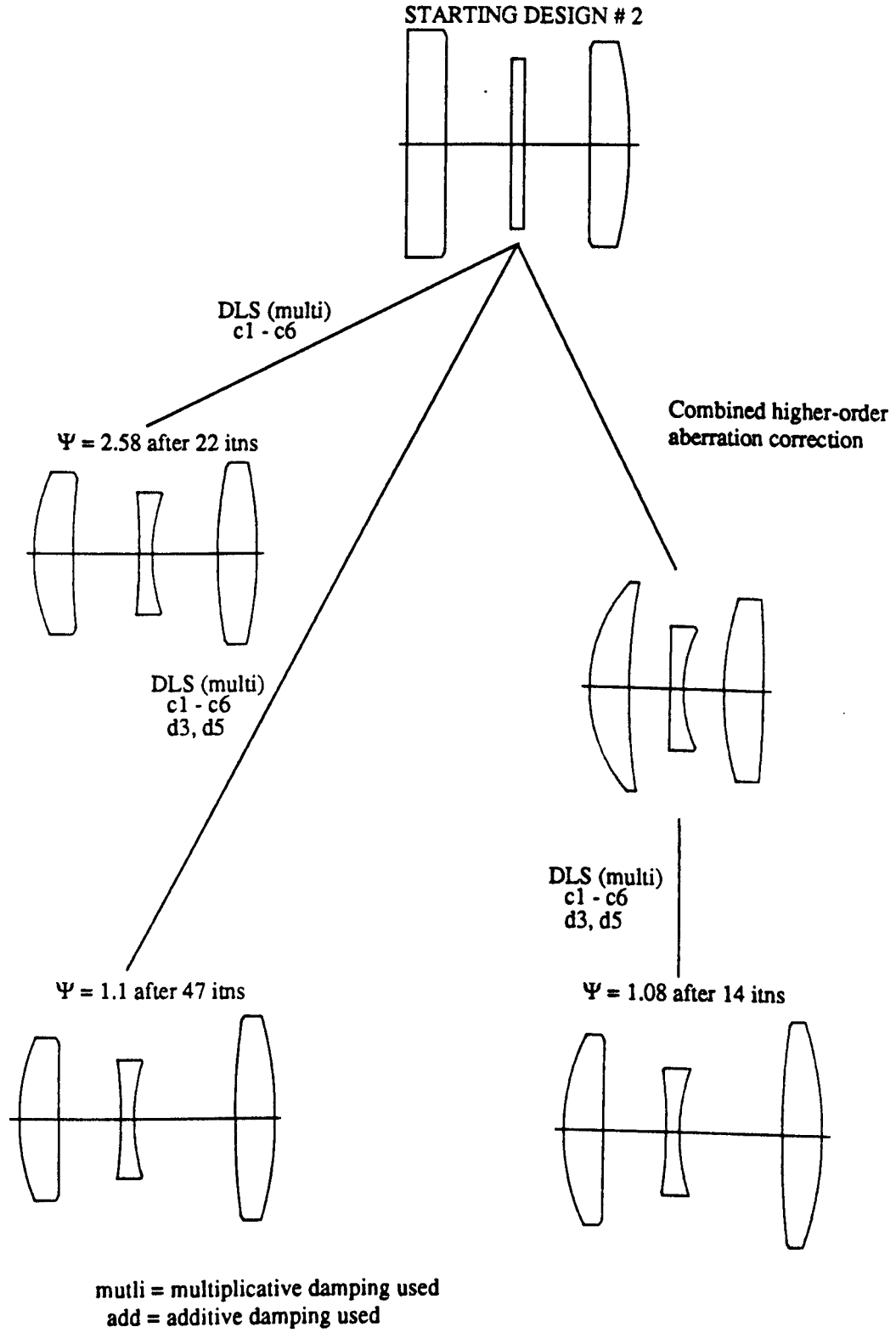


Figure 4.8 Diagrammatic representation of the lenses resulting from the corrections performed on triplet starting desing # 2 (see text).

This correction produces a design of the Cooke triplet form but the air spaces about the stop have been reduced symmetrically. (Grouped higher order aberration

correction of starting design # 1 in the previous section only reduced the first air space to less than 1 mm). Subsequent DLS optimization of this design with all curvatures and the two air spaces variable finds the asymmetric triplet solution with $\Psi = 1.1$. This design is identical with that achieved by DLS optimization of the starting design. However, optimization of the higher order corrected design takes only 14 iterations compared with the 47 originally required by DLS (fig. 4.6).

4.4.3 Experiments with a 50mm f/4 triplet - starting design # 3

The third starting design tested for the 50mm f/4 triplet was similar to starting design # 2 but the air spaces have been reduced from 5mm to 1mm (Appendix D, TL4.3). As a result this lens is a long way from the final Cooke triplet form and should present significant problems to an optimization program. The DLS run on this starting design with all curvatures and the two air spaces variable was successful and managed to extricate itself from at least one local minimum solution (see fig. 4.9). The final merit function after 52 iterations was $\Psi = 1.19$ and the lens form was an asymmetric triplet but with the central negative element closer to the rear positive lens (fig. 4.10). This was the opposite asymmetry to the designs found in §4.4.1 and §4.4.2. Its aberrational performance was slightly inferior due to larger edge of aperture aberrations.

Combined higher order correction on this design failed at the first iteration since the changes requested were so large that the curvatures were all frozen at their boundary values of 0.3. However, grouped sequential correction achieved a significant reduction in the mixed higher order aberrations: two groups were used based on the row rates of change as described in the previous section. The first group, consisting of M_I, M_{IV} and M_{VI} achieved 9/10ths of the requested change (target values, 0.05) and a second group consisting of all the mixed higher order aberrations with the exception of M_{II} achieved 4/10ths of the requested changes. No further higher order aberration correction was possible and the resulting values for the mixed higher order aberrations are given in Table 4.6.

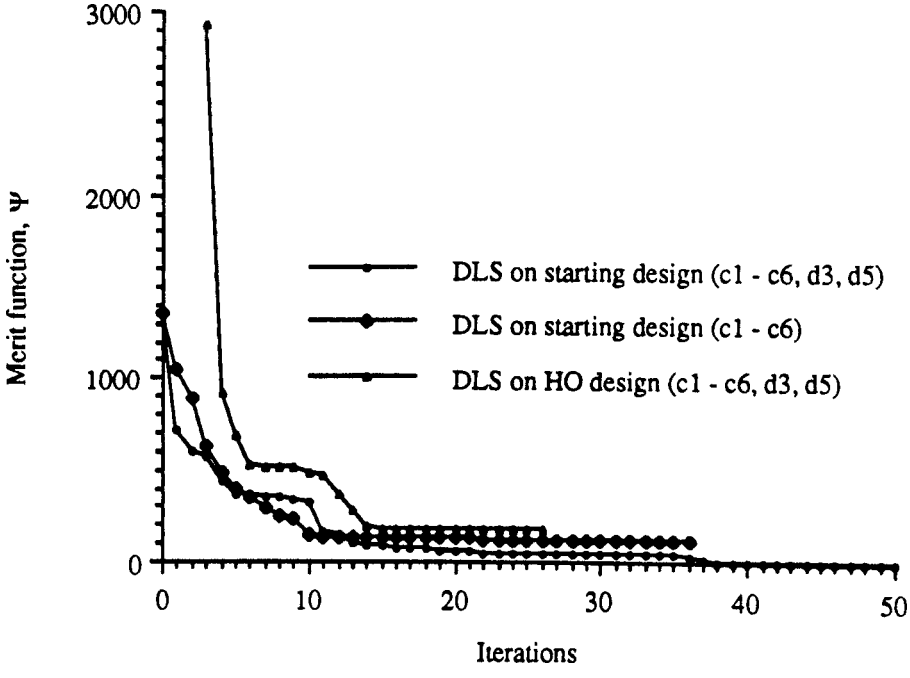


Figure 4.9 DLS optimization runs on starting design # 3 and higher order corrected designs.

Aberration	Start	pc4-10
M_I	-0.01603	0.03127
M_{II}	0.16693	-0.83954
M_{III}	1.15330	-0.11556
M_{IV}	0.55331	0.08112
M_V	-0.13212	0.05018
M_{VI}	-0.58785	0.00983
M_{VII}	-1.30831	0.09887

Table 4.6 Mixed Higher order Transverse Ray Aberrations both before ('start') and after ('pc4-10') grouped higher order aberration correction of an f/4 triplet (starting design # 3).

The resulting design is given in Figure 4.9 from which it can be seen that the air spaces have remained at their starting values of 1mm and the lens has some very strong curvatures to achieve the aberration correction. DLS optimization of this design stopped after 23 iterations with a merit function of $\Psi = 173.2$. The design is tending towards a double Gauss objective form with very steep curvatures about

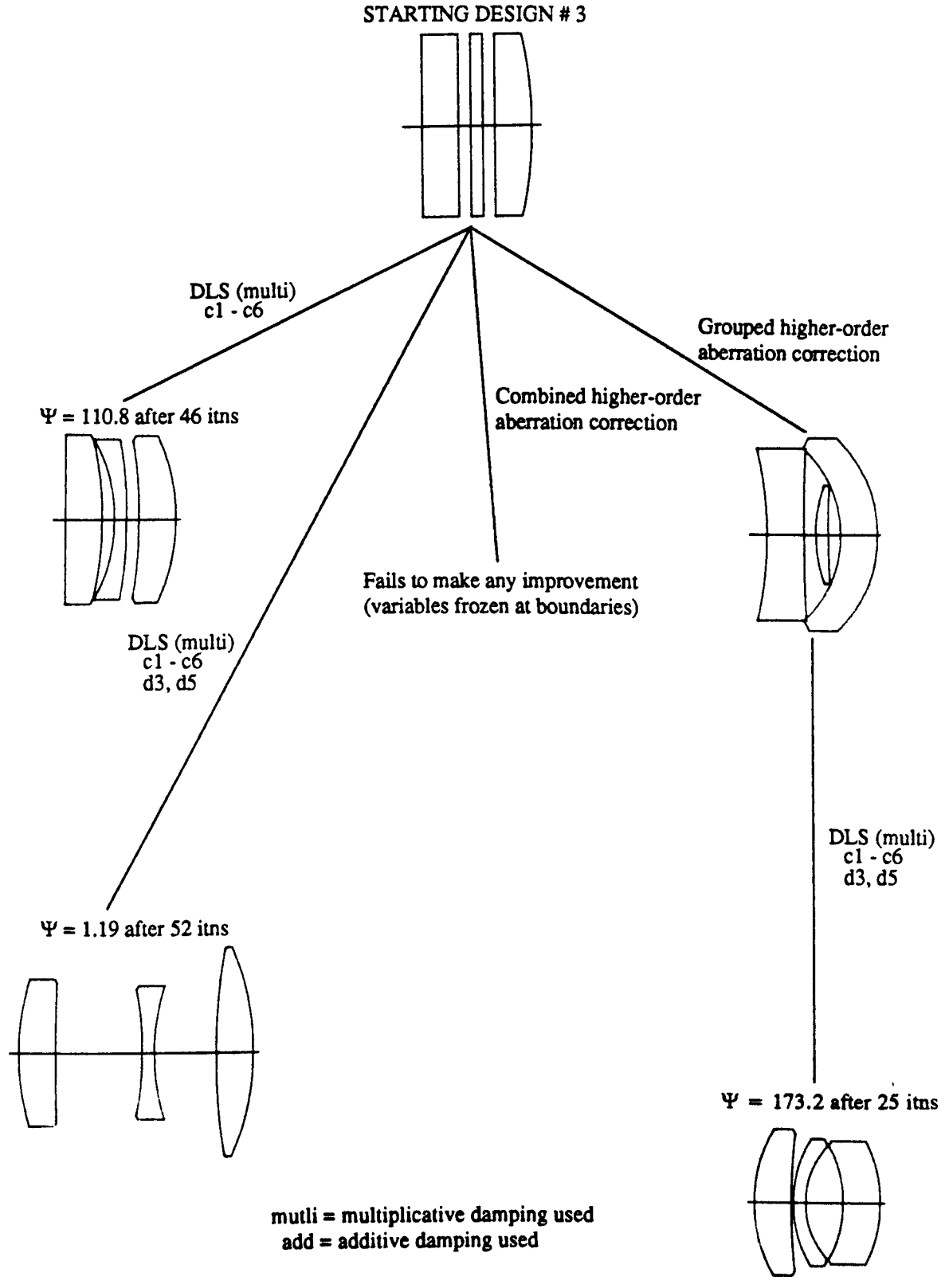


Figure 4.10 Diagrammatic representation of the lenses resulting from the corrections performed on triplet starting design # 3 (see text).

the stop. Aberrationally it is poor and suffers predominantly from longitudinal colour and oblique spherical aberration.

4.4.4 Experiments with an 90mm f/2 Petzval Lens - starting design # 4

The f/4 triplet used in the previous sections does not suffer from very large mixed higher order aberrations due to its moderate aperture. Therefore, a higher aperture lens was chosen for further testing of mixed higher order aberration correction. The lens chosen was an f/2 Petzval lens with a field flattener (see fig. 4.11). Its effective focal length was 90mm, the field ± 15 degrees and the glasses were Schott LF5/SF3 and BAF50/SF3 in the two doublets and SF3 for the field flattener. The doublets in the starting design (# 4) had bi-convex positive elements and plano-concave negative elements. The field flattener was plano-concave. The three air spaces were each 30mm although the assumption has been made that these are not significant variables and that they are close to their optimum values (Appendix D, TL4.4).

DLS optimization of this starting design had few problems and with all curvatures variable, converged to a design with a merit function of $\Psi = 18.85$ in 13 iterations (see fig. 4.12).

Combined higher order aberration correction was successful in reducing all mixed higher order aberrations to within the specified tolerance of their target values (0.05). Table 4.7 shows the extent to which the values have been reduced.

The lens resulting from this correction of the mixed higher order aberrations is shown in Figure 4.11 and the most noticeable change is the increased power of the positive elements of the doublets and the consequent increase in curvature at the doublet interfaces. DLS optimization of this design (fig. 4.12) converged to an equivalent solution to that obtained by applying DLS to the original starting design with a merit function of $\Psi = 19.17$ after 17 iterations.

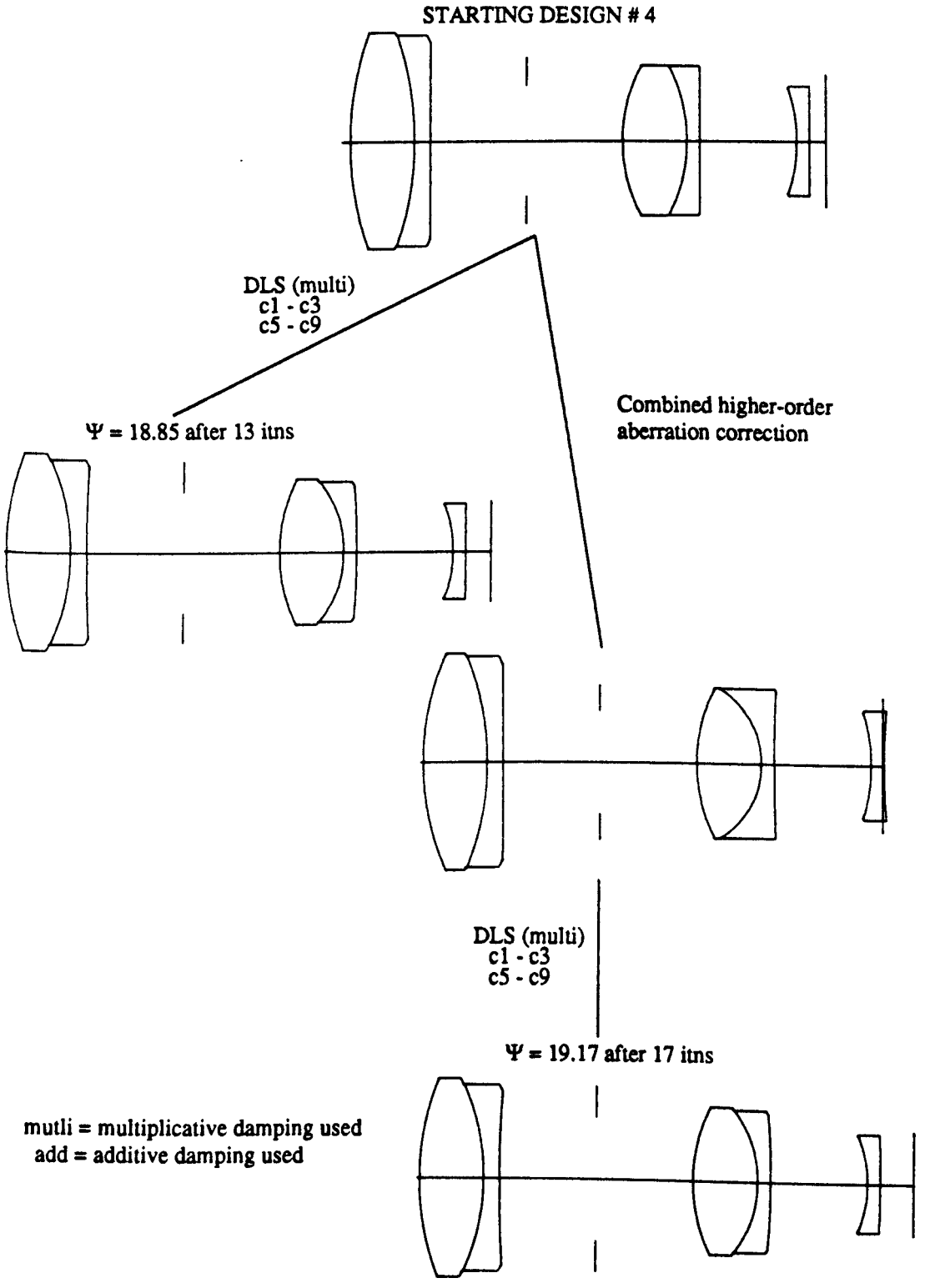


Figure 4.11 Diagrammatic representation of the lenses resulting from the corrections performed on the Petzval lens starting design # 4 (see text).

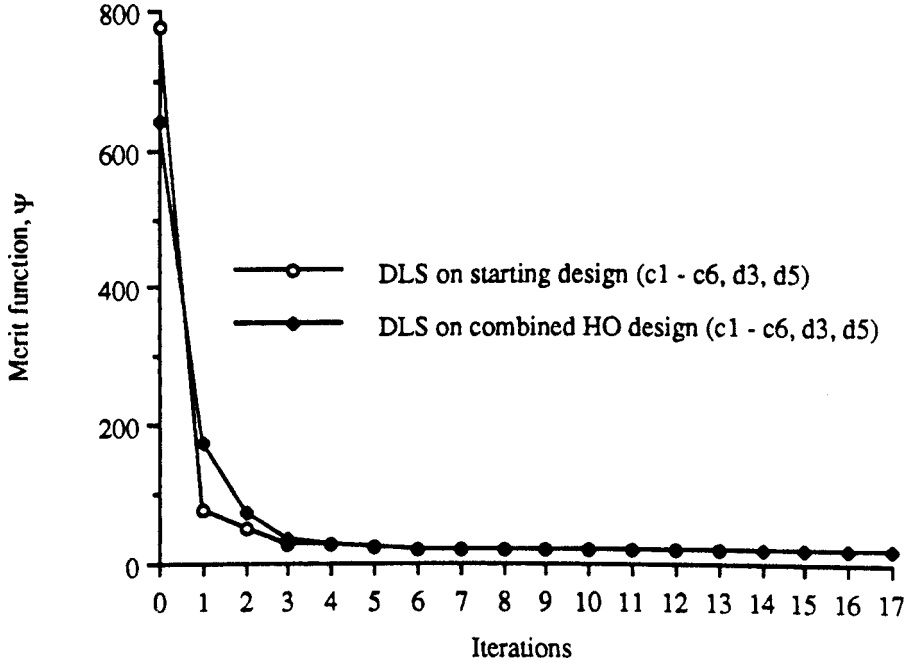


Figure 4.12 DLS optimization runs on starting design # 4 and higher order corrected designs.

Aberration	Start	Corrected
M_I	0.03309	0.05010
M_{II}	0.19132	0.04977
M_{III}	1.54202	0.04986
M_{IV}	0.63564	0.04968
M_V	0.06607	0.05032
M_{VI}	-0.61486	0.04963
M_{VII}	-1.49108	0.46045

Table 4.7 Mixed Higher order Transverse Ray Aberrations both before ('start') and after ('corrected') grouped higher order aberration correction of an f/2 Petzval lens (starting design # 4).

For this example, no benefit has been gained from achieving some degree of correction of the higher order aberrations either in terms of efficiency or convergence to a different solution. In fact the Petzval lens solution was found to be very stable with a large capture region using the DLS algorithm. Attempts to force different

solutions by collapsing the air spaces surrounding the stop and varying the aperture and field requirements failed to find any different solutions until the aperture was reduced to f/10 with a field of ±40 degrees when a double Gauss form was found.

4.4.5 Experiments with a 75mm f/1 IR triplet - starting design # 5

The aperture was increased to f/1 for the final example which was an infra-red triplet with an EFL of 75mm and field of ±5.65 degrees. The materials chosen for the lenses were Germanium for the outer elements and NaCl for the central element. The starting design (# 5) is shown in figure 4.13 and is a somewhat eccentric looking Cooke triplet with the stop at the first surface (Appendix D, TL4.5). DLS optimization of this design with all curvatures and the two air spaces variable converged in 22 iterations to a design which is close to the optimum and has a merit function of $\Psi = 424.75$ ($\Psi \approx 63$ without edge violations). Both additive and multiplicative damping, indicated by 'add' and 'multi' in figure 4.13, gave an identical solution although multiplicative damping took 29 iterations before stopping.

Aberration	Row rate of change
M_I	7309.81
M_{II}	2761.08
M_{III}	24.95
M_{IV}	15.39
M_V	2.78
M_{VI}	58.50
M_{VII}	1867.90

Table 4.8 RRC's during grouped higher order aberration correction of starting design # 5.

Combined correction of all seven mixed higher order aberrations could do no better than achieve 1/10th of the requested aberration change and so grouped sequential correction was attempted. Following a similar approach to that taken in the previous sections, the four slowest aberrations, M_{III}, M_{IV}, M_V and M_{VI}, were

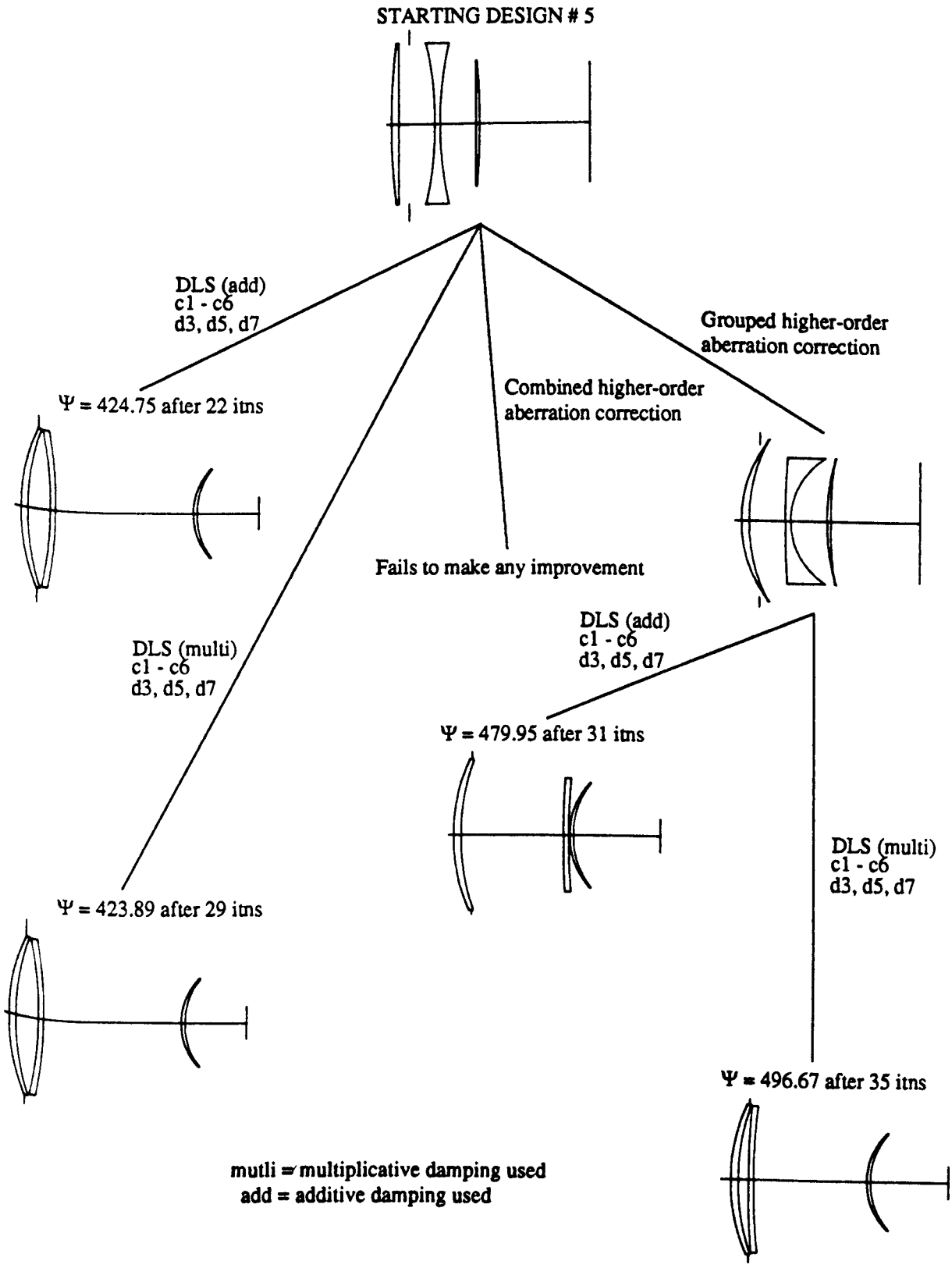


Figure 4.13 Diagrammatic representation of the lenses resulting from the corrections performed on the IR triplet starting desing # 5 (see text).

first targeted and reduced to their required level of 0.05. The row rates of change for this example (Table 4.8) were very disparate and so the very fast aberrations (M_I, M_{II} and M_{VII}) were targeted separately on the assumption that the changes in the system parameters would be relatively small and therefore not disrupt the correction previously achieved for M_{III}, M_{IV}, M_V and M_{VI}.

Aberration	Start	pc6-10
M_I	-1.08251	0.18217
M_{II}	0.56145	0.54456
M_{III}	0.82794	0.28739
M_{IV}	0.23602	-0.02507
M_V	-0.01498	-0.01488
M_{VI}	-0.24719	0.07250
M_{VII}	-0.92150	-0.39050

Table 4.9 Mixed Higher order Transverse Ray Aberrations both before ('start') and after ('pc6-10') grouped higher order aberration correction of an f/2 Petzval lens (starting design # 4).

Correction of the second group only achieved 6/10ths of the requested aberration change and the resulting values for the mixed higher order aberrations after this correction are given in Table 4.9. The form of the lens resulting from this correction is show in figure 4.13 and it is tending towards three meniscus components all with positive curvatures.

DLS optimization of this design converges to two different solutions depending on the method of damping used. If multiplicative damping is used, the final design has an air spaced meniscus doublet for the first group and a strong meniscus lens which acts as a field flattener in the second group. However, if additive damping is used then the resulting design has a singlet in the first group and a combination of a near plane parallel plate and a strong meniscus lens in the second group. Both these design are aberrationally inferior to the original DLS correction. The number of iterations taken for convergence (fig. 4.14) is also greater although edge violations have a significant influence in this case.

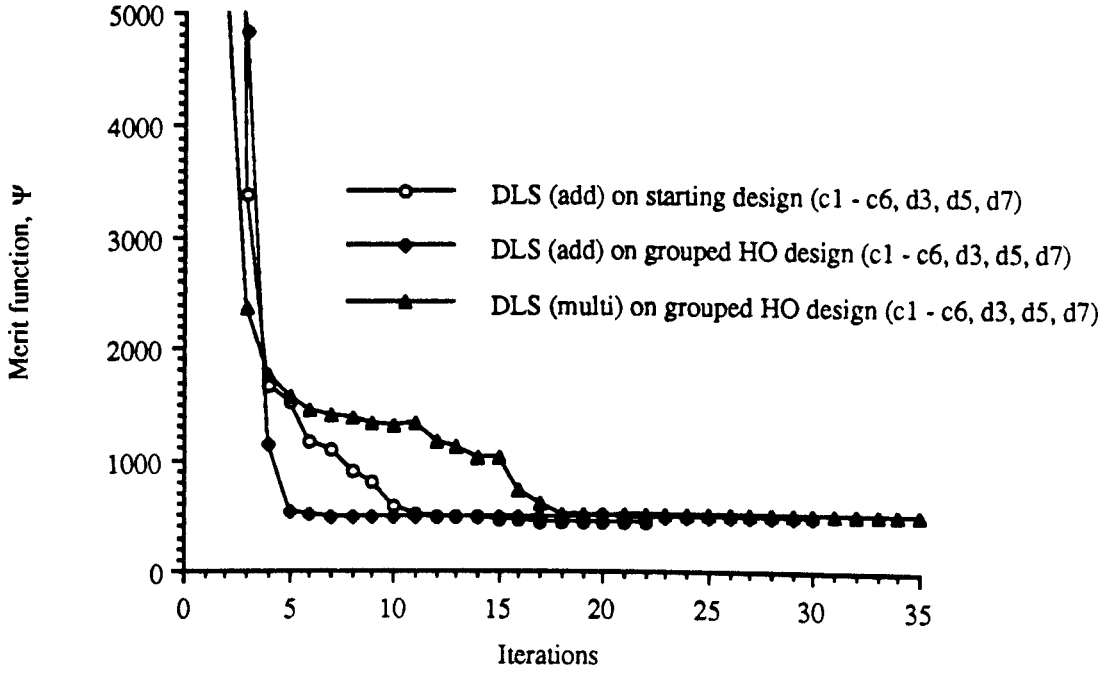


Figure 4.14 DLS optimization runs on starting design # 5 and higher order corrected designs.

4.5 References

[4.1] Rogers, P.J., *Private Communication*, (1991).

[4.2] Cox, A., *A System of Optical Design*, Focal Press, London (1964).

[4.3] Welford, W.T. *Aberrations of Optical Systems*, Adam Hilger Ltd. (1986), p265.

[4.4] *ibid*, p 187.

CHAPTER 5

The Fine Correction of Optical Systems

5.1 Introduction

Conrady [5.1] stated that the fundamental principle of the treatment of image forming optical systems is that "the optical paths from object to image should all be equal to within a fraction of a wavelength". In the optimization of an optical system this task can be nominally split into two parts: the reduction of aberration values and the adjustment of the relationship between the various aberrations to achieve an overall optimum state of correction. The latter process, where higher order aberrations which are irreducible in some ill-defined sense, are balanced by third order aberrations, is generally referred to as aberration balancing. Aberration balancing is usually performed during the final stages of the design process when it is desirable to optimize some meaningful measure of image quality. The aim is not necessarily to reduce the magnitudes of the aberrations further, (although this would be beneficial), but to make small changes to the system which cause the aberration coefficients to have a nearly equal but opposite effect, the result of which would be a very low residual aberration.

Many different approaches have been applied to this problem and most optimization programs have some aberration balancing potential. In fact, an optimization program that didn't perform some aberration balancing would not be particularly useful. The damped least squares method, for example, usually attempts to min-

imize a merit function defined as the sum of the squares of the transverse ray aberrations which have been determined by a system raytrace. The selection of rays that are traced helpsto determine the aberration balance. In addition, the aberrations in the merit function are usually multiplied by a weighting factor which affects the aberration balance with respect to the field coordinate. Although this approach has yielded satisfactory results, it can lead to different 'best' forms of correction dependent upon the choice of aberrations (the sampling points) and the weighting factors.

An alternative approach to the problem is to define a new set of aberration coefficients which, when zero, balance the aberrations in the Hamiltonian expansion. Several author s [5.2, 5.3] have described the expansion of the classical aberrations in terms of orthogonal polynomials which have an inherent ability to balance aberrations. For instance, the Zernike polynomial $R_6^0 = 20r^6 - 30r^4 + 12r^2 - 1$, where r is the relative aperture coordinate, can be recognized as a combination of fifth order spherical aberration, third order spherical aberration and defocus [5.4]. The significant feature is that the coefficients are in exactly the correct ratio to maximize the Strehl ratio (see §5.4).

It is generally recognized [5.5–5.7] that different criteria of image quality are required for automatic correction programs during the final stages of design than those used during the early stages of optimization. For example, the merit function based on the sum of the squares of the transverse ray aberrations becomes inadequate when the system approaches the diffraction limit. In these situations, suitable merit functions that can be used include RMS spot size [5.8, 5.9], wavefront variance [5.6, 5.10], the variance of a wave aberration difference function [5.11] and direct optimization of the spatial frequency response (OTF) [5.12, 5.13]. Most of these techniques are limited to well corrected systems, although Tatian [5.14] has described a balancing criterion for non-diffraction limited systems. Comparisons of several of the optimization criteria have been carried out by Hopkins [5.6], Meiron [5.7] and Woodruff [5.15].

In this chapter, values of the third order aberration coefficients are derived which balance the fifth order aberrations to maximize the Strehl ratio (§5.4). This, in turn, maximizes the volume under the spatial frequency response curves, a result which follows from the inverse Fourier transform relationship between the irradiance distribution and the two dimensional optical transfer function. The third order coefficients are used as target values for the quadratic adaptive correction algorithm described in chapter 3 and applied to several test lenses (§5.5).

5.2 Wavefront Aberration Approximation to Fifth order

Wavefront aberration polynomial coefficients of higher order than third may be computed using the methods of Buchdahl [5.16] or by polynomial fitting to the exact raytrace-computed wavefront aberrations [5.17]. Knowledge of these higher order coefficients, which have lower rates of change than the third order coefficients, allow targets to be set for the Seidel aberrations at any stage of the design process. However, it is only when the design is close to its optimum correction balance that this technique is useful, because it is then that the relative magnitudes of the higher order terms is large enough for them to play a significant part in determining the image quality.

The wavefront aberration can be approximated by the wavefront polynomial containing all terms up to fifth order together with seventh order spherical aberration for most systems of moderate aperture and field. In this section values for the fifth order coefficients are derived from a knowledge of the Seidel sums and nine values of the finite wavefront aberration calculated during a system raytrace (fig. 5.1).

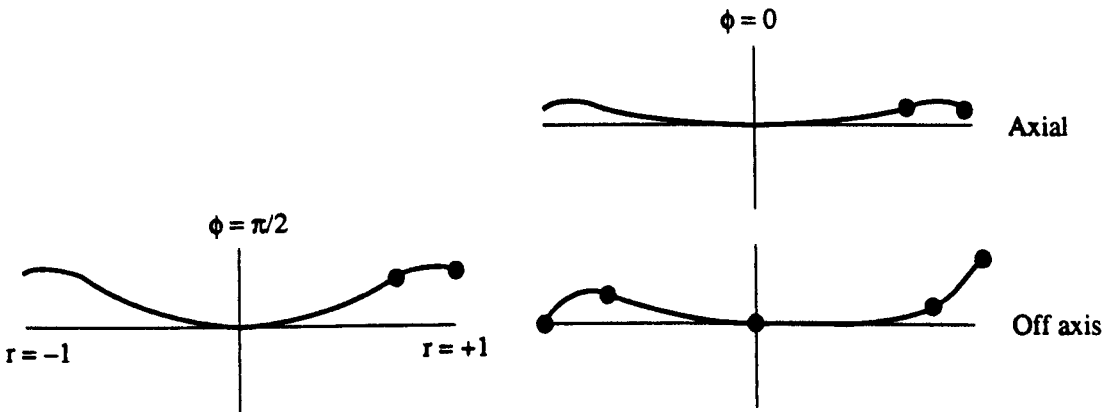


Figure 5.1. Finite wavefront aberration values required to approximate the wavefronts to a fifth order expansion.

The wavefront aberration, $W(r, \phi, \varrho)$, expressed in polar coordinates (r, ϕ) and relative field position ϱ, contains both even and odd terms with respect to the aperture coordinate. The even terms are due to spherical aberration, astigmatism and field curvature, whereas the odd terms are due to the coma-like components of the total aberration plus distortion and transverse chromatic aberration. The even component of the monochromatic wavefront aberration, W_e, is therefore given by

$$W_e = {}_0w_{20}r^2 + {}_0w_{40}r^4 + {}_0w_{60}r^6 + {}_0w_{80}r^8 + {}_2w_{40}\varrho^2r^4 + {}_2w_{42}\varrho^2r^4\cos^2\phi + {}_2w_{20}\varrho^2r^2 + {}_2w_{22}\varrho^2r^2\cos^2\phi + {}_4w_{20}\varrho^4r^2 + {}_4w_{22}\varrho^4r^2\cos^2\phi \qquad (5.1)$$

and the odd component by

$$W_o = {}_1w_{31}\varrho r^3\cos\phi + {}_1w_{51}\varrho r^5\cos\phi + {}_3w_{31}\varrho^3r^3\cos\phi + {}_3w_{33}\varrho^3r^3\cos^3\phi. \qquad (5.2)$$

Although there are ten terms in equation (5.1), six are dependent on field and three on pupil azimuth. As a result, it is possible to find the wavefront aberration polynomial coefficients by considering the wavefront aberrations calculated for the axial field position and one off-axis field position for both the meridian and sagittal sections. During a system raytrace, values for the total wavefront aberration are assumed to be computed for both the margin ($r = 1$) and zone ($r = 1/\sqrt{2}$) of the aperture in both the meridian and sagittal sections for each field position. Therefore, values of the finite wavefront aberration $W(r, \phi, \varrho)$ are obtained for

$$\phi = 0, \pi/2 \begin{cases} r = \pm 1 \\ r = \pm 1/\sqrt{2} \end{cases} \qquad (5.3)$$

The values of $W(r, \phi, \varrho)$ for $r < 0$ both in the sagittal section and for the axial field position are derived from symmetry.

The third order terms in the wavefront polynomial expansion can be determined from the values of the Seidel sums which have been calculated from an appropriately defined paraxial ray trace of the system. The five third order monochromatic aberration coefficients, as they appear in the Hamiltonian wavefront expansion, are given in table 5.1 when expressed in terms of the polar coordinates (r, ϕ, ϱ).

Aberration	Third-order term
Spherical aberration	${}_0w_{40}r^4$
Coma	${}_1w_{31}\varrho r^3\cos\phi$
Astigmatism and field curvature	${}_2w_{22}\varrho^2r^2\cos^2\phi + {}_2w_{20}\varrho^2r^2$
Distortion	${}_3w_{11}\varrho^3r\cos\phi$

Table 5.1 Third order monochromatic wavefront aberrations.

Their relationship to the Seidel sums can be found by direct comparison of the expressions in table 5.1 with, for example, equation (8.42) in reference [5.18]:

$$ {}_0w_{40} = \tfrac{1}{8}S_I \tag{5.4} $$

$$ {}_1w_{31} = \tfrac{1}{2}S_{II} \tag{5.5} $$

$$ {}_2w_{20} = \tfrac{1}{4}(S_{III} + S_{IV}) \tag{5.6} $$

$$ {}_2w_{22} = \tfrac{1}{2}S_{III} \tag{5.7} $$

$$ {}_3w_{11} = \tfrac{1}{2}S_V. \tag{5.8} $$

The value of ${}_0w_{20}$ is found by deriving its relation to the longitudinal shift in the image plane, Δz, which has been specified by the designer or an optimization program. Figure 5.2 shows the geometry for the axial reference sphere and a defocussed wavefront.

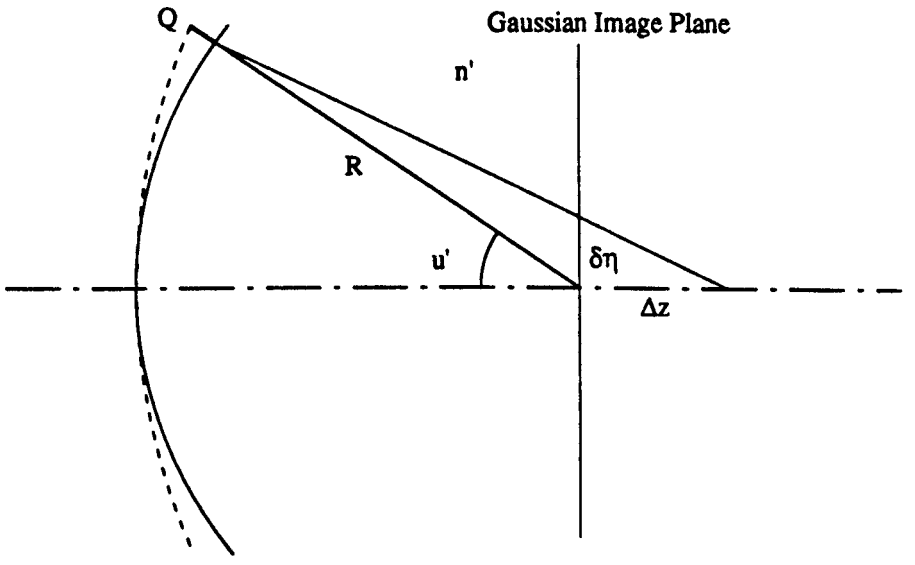


Figure 5.2. Wavefront aberration component of longitudinal defocus.

The wavefront aberration for defocus is given by

$$W = {}_0w_{20}\left(\frac{x_p^2 + y_p^2}{h_p^2}\right) \tag{5.9}$$

where x_p and y_p are the cartesian coordinates in the exit pupil and h_p the height of the paraxial marginal ray at the pupil. Differentiating equation (5.9) gives the transverse ray aberration component of defocus, which after substituting for the radius of the reference sphere $R = -h_p/u'$ becomes

$$\delta\eta = \frac{2\,{}_0w_{20}}{n'u'}r \tag{5.10}$$

where r is the normalized height in the exit pupil. The longitudinal component of defocus Δz is therefore given by $\delta\eta/u'$ and hence

$${}_0w_{20} = \tfrac{1}{2}n'u'^2\Delta z \tag{5.10}$$

for the marginal ray when $r = 1$.

The axial spherical aberration coefficients can be derived from the wavefront aberrations at the axial field position ($\varrho = 0$). Inspection of equation (5.1) shows that only these first four terms are non-zero and therefore the following two equations can be written for the finite wavefront aberration calculated at the margin and the zone of the aperture respectively:

$$W(1,0,0) = {}_0w_{20} + {}_0w_{40} + {}_0w_{60} + {}_0w_{80} \tag{5.12}$$

$$W(1/\sqrt{2},0,0) = \tfrac{1}{2}\,{}_0w_{20} + \tfrac{1}{4}\,{}_0w_{40} + \tfrac{1}{8}\,{}_0w_{60} + \tfrac{1}{16}\,{}_0w_{80}. \tag{5.13}$$

The coefficients for defocus and third order spherical aberration are known from equations (5.4) and (5.11) and therefore solving for ${}_0w_{60}$ and ${}_0w_{80}$ gives that

$$ {}_0w_{60} = 16W(1/\sqrt{(2)},0,0) - W(1,0,0) - 7\,{}_0w_{20} - 3\,{}_0w_{40} \tag{5.14} $$

$$ {}_0w_{80} = 2W(1,0,0) - 16W(1/\sqrt{(2)},0,0) + 6\,{}_0w_{20} + 2\,{}_0w_{40}. \tag{5.15} $$

In the sagittal section, $\phi = \pi/2$ and all the terms in equations (5.1) and (5.2) containing powers of $\cos\phi$ are zero. As a result the total wavefront aberration, $W(r,\pi/2,1)$, for the full semi-field position ($\varrho = 1$) is given by

$$ \begin{aligned} W(r,\pi/2,1) = {}& {}_0w_{20}r^2 + {}_0w_{40}r^4 + {}_0w_{60}r^6 + {}_0w_{80}r^8 \\ & + {}_2w_{40}r^4 + {}_2w_{20}r^2 + {}_4w_{20}r^2. \end{aligned} \tag{5.16} $$

At the margin of the aperture (r=1) the wavefront aberration is therefore

$$ W(1,\pi/2,1) = {}_0w_{20} + {}_0w_{40} + {}_0w_{60} + {}_0w_{80} + {}_2w_{40} + {}_2w_{20} + {}_4w_{20} \tag{5.17} $$

and at the zone of the aperture ($r = 1/\sqrt{2}$)

$$ \begin{aligned} W(1/\sqrt{2},\pi/2,1) = {}& \tfrac{1}{2}\,{}_0w_{20} + \tfrac{1}{4}\,{}_0w_{40} + \tfrac{1}{8}\,{}_0w_{60} + \tfrac{1}{16}\,{}_0w_{80} \\ & + \tfrac{1}{4}\,{}_2w_{40} + \tfrac{1}{2}\,{}_2w_{20} + \tfrac{1}{2}\,{}_4w_{20} \end{aligned} \tag{5.18} $$

Equations (5.17) and (5.18) can be solved for ${}_2w_{40}$ and ${}_4w_{20}$ to give that

$$ {}_2w_{40} = 2W(1,\pi/2,1) - 4W(1/\sqrt{2},\pi/2,1) - {}_0w_{40} - \tfrac{3}{2}\,{}_0w_{60} - \tfrac{7}{4}\,{}_0w_{80} \tag{5.19} $$

and

$$ {}_4w_{20} = 4W(1/\sqrt{2},\pi/2,1) - W(1,\pi/2,1) - {}_0w_{20} + \tfrac{1}{2}\,{}_0w_{60} + \tfrac{3}{4}\,{}_0w_{80} - {}_2w_{20} \tag{5.20} $$

since ${}_2w_{20}$ is known from equation (5.6).

Alternatively, it is possible to write down a pair of equations which can be solved for ${}_2w_{40}$ and ${}_4w_{20}$ at the zone of the field, or a pair of equations can be written using both off-axis field positions and only the zonal aperture rays. This raises the question as to which field position should be used for computing the aberration

coefficients. For example, it is probable that the designer will want a higher image quality at the zone of the field than at the field margin and therefore it would be better to use the zone of the field for computing the higher order aberrations. The selection of fitting points can greatly affect the approximation for designs with different aberrational characteristics; this is examined in more detail in §5.3 when numerical and graphical results are presented for several test lenses.

In the meridian section ($\phi = 0$) and at the full semi-field position, the approximating wavefront aberration polynomial contains all terms in equations (5.1) and (5.2). Therefore

$$W(r,0,1) = W_e + W_o \tag{5.21}$$

where powers of ϱ can be removed since they are all unity. The total wavefront aberration, $W(r,0,1)$, for $r = \pm 1$ and $r = \pm 1/\sqrt{2}$ has been computed from a system raytrace, therefore at the upper and lower margins of the aperture ($r = \pm 1$)

$$\begin{aligned} W(1,0,1) = {}& {}_0w_{20} + {}_0w_{40} + {}_0w_{60} + {}_0w_{80} + {}_2w_{40} + {}_2w_{42} + {}_1w_{31} + {}_1w_{51} \\ & + {}_3w_{31} + {}_3w_{33} + {}_2w_{20} + {}_2w_{22} + {}_4w_{20} + {}_4w_{22} \end{aligned} \tag{5.22}$$

$$\begin{aligned} W(-1,0,1) = {}& {}_0w_{20} + {}_0w_{40} + {}_0w_{60} + {}_0w_{80} + {}_2w_{40} + {}_2w_{42} - {}_1w_{31} - {}_1w_{51} \\ & - {}_3w_{31} - {}_3w_{33} + {}_2w_{20} + {}_2w_{22} + {}_4w_{20} + {}_4w_{22}. \end{aligned} \tag{5.23}$$

Similarly for the zone of the aperture ($r = \pm 1/\sqrt{2}$)

$$\begin{aligned} W(1/\sqrt{2},0,1) = {}& +\tfrac{1}{2}\,{}_0w_{20} + \tfrac{1}{4}\,{}_0w_{40} + \tfrac{1}{8}\,{}_0w_{60} + \tfrac{1}{16}\,{}_0w_{80} + \tfrac{1}{4}\,{}_2w_{40} + \tfrac{1}{4}\,{}_2w_{42} \\ & + \tfrac{1}{2\sqrt{2}}\,{}_1w_{31} + \tfrac{1}{4\sqrt{2}}\,{}_1w_{51} + \tfrac{1}{2\sqrt{2}}\,{}_3w_{31} + \tfrac{1}{2\sqrt{2}}\,{}_3w_{33} + \tfrac{1}{2}\,{}_2w_{20} \\ & + \tfrac{1}{2}\,{}_2w_{22} + \tfrac{1}{2}\,{}_4w_{20} + \tfrac{1}{2}\,{}_4w_{22} \end{aligned} \tag{5.24}$$

$$\begin{aligned} W(-1/\sqrt{2},0,1) = {}& +\tfrac{1}{2}\,{}_0w_{20} + \tfrac{1}{4}\,{}_0w_{40} + \tfrac{1}{8}\,{}_0w_{60} + \tfrac{1}{16}\,{}_0w_{80} + \tfrac{1}{4}\,{}_2w_{40} + \tfrac{1}{4}\,{}_2w_{42} \\ & - \tfrac{1}{2\sqrt{2}}\,{}_1w_{31} - \tfrac{1}{4\sqrt{2}}\,{}_1w_{51} - \tfrac{1}{2\sqrt{2}}\,{}_3w_{31} - \tfrac{1}{2\sqrt{2}}\,{}_3w_{33} + \tfrac{1}{2}\,{}_2w_{20} \\ & + \tfrac{1}{2}\,{}_2w_{22} + \tfrac{1}{2}\,{}_4w_{20} + \tfrac{1}{2}\,{}_4w_{22}. \end{aligned} \tag{5.25}$$

The above equations form an underdetermined linear system since there are five unknown coefficients and only four equations. However, the elliptical coma terms,

${}_3w_{31}\varrho^3r^3\cos\phi$ and ${}_3w_{33}\varrho^3r^3\cos^3\phi$, both have identical values in the meridian section and zero contribution in the sagittal section. Figure 5.3 shows the variation of ${}_3w_{31}$ and ${}_3w_{33}$ with azimuth and demonstrates that different contributions from ${}_3w_{31}$ and ${}_3w_{33}$ can only be distinguished for intermediate azimuths between the meridian section ($\phi = 0$) and the sagittal section ($\phi = \pi/2$). Consequently, it is necessary to consider the sum of the elliptical coma terms if the wavefront aberrations are not calculated at an intermediate azimuth.

By defining $\gamma = {}_3w_{31}\varrho^3r^3\cos\phi + {}_3w_{33}\varrho^3r^3\cos^3\phi$, the above equations can be solved for the four remaining variables as follows: Taking the sum and the difference of equations (5.22) and (5.23), the following two equations are obtained:

$$\begin{aligned} {}_2w_{42} + {}_4w_{22} = {} & \tfrac{1}{2}\left(W(1,0,1) + W(-1,0,1)\right) - {}_0w_{20} - {}_0w_{40} \\ & - {}_0w_{60} - {}_0w_{80} - {}_2w_{40} - {}_2w_{20} - {}_2w_{22} - {}_4w_{20} \end{aligned} \tag{5.26}$$

$${}_1w_{51} + \gamma = \tfrac{1}{2}\left(W(1,0,1) - W(-1,0,1)\right) - {}_1w_{31}. \tag{5.27}$$

Similarly for equations (5.24) and (5.25)

$$\begin{aligned} {}_2w_{42} + {}_4w_{22} = {} & \left(W(1/\sqrt{2},0,1) + W(-1/\sqrt{2},0,1)\right) - {}_0w_{20} - \tfrac{1}{2}\,{}_0w_{40} - \tfrac{1}{4}\,{}_0w_{60} \\ & - \tfrac{1}{8}\,{}_0w_{80} - \tfrac{1}{2}\,{}_2w_{40} - {}_2w_{20} - {}_2w_{22} - {}_4w_{20} \end{aligned} \tag{5.28}$$

$$\tfrac{1}{2}\,{}_1w_{51} + \gamma = \sqrt{2}\left(W(1/\sqrt{2},0,1) - W(-1/\sqrt{2},0,1)\right) - {}_1w_{31} \tag{5.29}$$

Replacing the right hand sides (RHS) of equations (5.26–5.29) with the constants A, B, C, and D respectively, (the RHS's all contain known coefficients), equations (5.26) and (5.28) can be solved for ${}_2w_{42}$ and ${}_4w_{22}$ to give

$${}_2w_{42} = 2(A - C) \tag{5.30}$$

$${}_4w_{22} = 2C - A \tag{5.31}$$

and equations (5.27) and (5.29) can be solved for ${}_1w_{51}$ and γ to give

$${}_1w_{51} = 2(B - D) \tag{5.32}$$

$$\gamma = 2D - B. \tag{5.33}$$

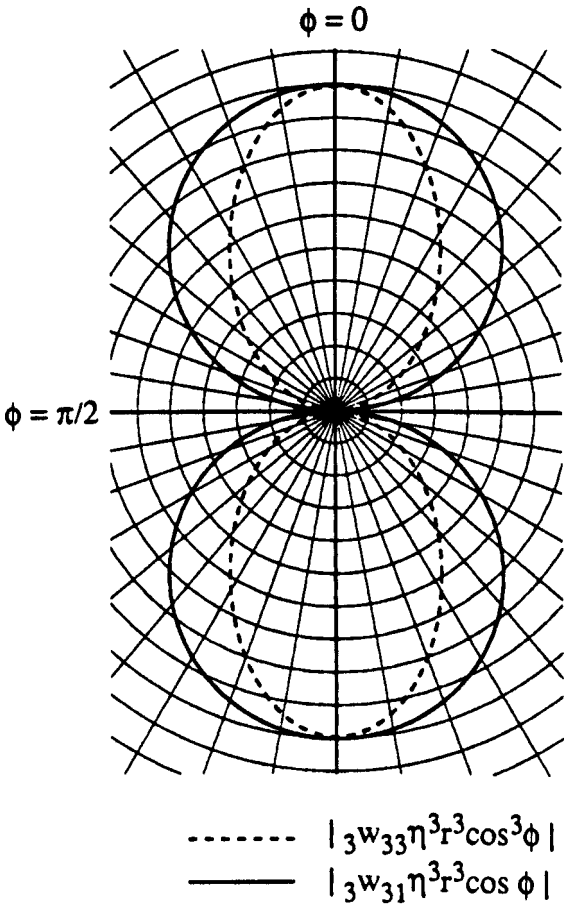


Figure 5.3. Polar diagram showing the variation of elliptical coma with azimuth.

Finally, an approximate value for fifth order distortion can be derived but in a slightly different manner to the results already presented. This is because most programs that calculate finite wavefront aberrations define a reference sphere centred on the point at which the finite principal ray strikes the image plane. Distortion is then either presented as numerical data during a system raytrace or as distortion curves plotted for the whole field. However, the transverse ray aberration component of distortion is readily available and this can be used to calculate the corresponding wavefront aberration coefficient.

The transverse ray aberration component of distortion is the difference between the height at which the paraxial principal ray and the finite principal ray strike the

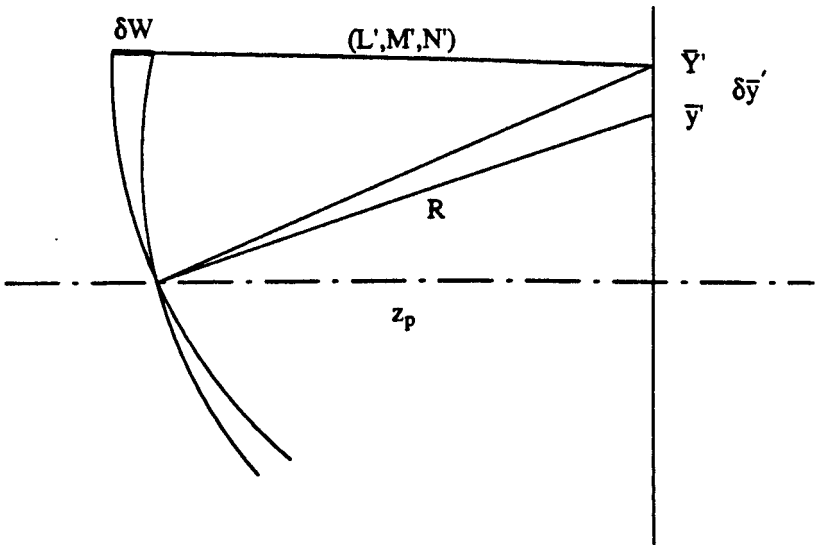


Figure 5.4. The relationship between distortion and wavefront aberration.

image plane (fig. 5.4),

$$\delta\bar{y}' = \bar{y}' - \bar{Y}'. \tag{5.34}$$

Distortion therefore produces a lateral shift in the centre of the reference sphere. The equation for the reference sphere which is centred on the height at which the paraxial principal ray strikes the image plane, $\bar{y}'$, is given by

$$x^2 + y^2 + z^2 - 2y\bar{y}' - 2zz_p = 0 \tag{5.35}$$

where z_p is the separation of the exit pupil from the image plane. Equation (5.35) can be differentiated to give

$$x\delta x + y\delta y + z\delta z = y\,\delta\bar{y}' + \bar{y}'\delta y + z_p\delta z. \tag{5.36}$$

which can be simplified using the following identities:

$$\delta x = -L\frac{\delta W}{n} \tag{5.37}$$

$$\delta y = -M\frac{\delta W}{n} \tag{5.38}$$

$$\delta z = -N\frac{\delta W}{n} \tag{5.39}$$

to give

$$\delta W = \frac{-n\,y\,\delta\bar{y}'}{Lx + M(y - \bar{y}') + N(z - z_p)}. \tag{5.40}$$

Finally, after substitution for the direction cosines and further simplification

$$\delta W = \frac{n}{R}\, y\, \delta\bar{y}' \tag{5.41}$$

which is the finite wavefront aberration for distortion. The approximate fifth order coefficient for distortion, ${}_5w_{11}$, can now be found;

$$\begin{aligned} {}_5w_{11} &= \delta W - {}_3w_{11} \\ &= \frac{n}{R}\, y\, \delta\bar{y}' - \tfrac{1}{2} S_V \end{aligned} \tag{5.42}$$

Table 5.2 summarises the results from this section.

Aberration	Approximate Value
${}_0w_{20}$	$\Delta z(\frac{1-N'}{N'})$
${}_0w_{60}$	$16W(1/\sqrt{(2)},0,0) - W(1,0,0) - 7\,{}_0w_{20} - 3\,{}_0w_{40}$
${}_0w_{80}$	$2W(1,0,0) - 16W(1/\sqrt{(2)},0,0) + 6\,{}_0w_{20} + 2\,{}_0w_{40}$
${}_2w_{40}$	$2W(1,\pi/2,1) - 4W(1/\sqrt{2},\pi/2,1) - {}_0w_{40} - \frac{3}{2}\,{}_0w_{60} - \frac{7}{4}\,{}_0w_{80}$
${}_4w_{20}$	$4W(1/\sqrt{2},\pi/2,1) - W(1,\pi/2,1) - {}_0w_{20} + \frac{1}{2}\,{}_0w_{60} + \frac{3}{4}\,{}_0w_{80} - {}_2w_{20}$
${}_2w_{42}$	$2(A-C)$
${}_4w_{22}$	$2C - A$
${}_1w_{51}$	$2(B-D)$
γ	$2D - B$
${}_5w_{11}$	$\frac{n}{R} y \delta\bar{y}' - \frac{1}{2} S_V$

Table 5.2 Approximate fifth order aberration coefficients.

5.3 Testing the Fifth Order Aberration Coefficients.

The results of the previous section are only approximate in that a complete description of the wavefront aberration of an optical system is only obtained if all the aberration coefficients are found to all orders. This is impractical and in all cases when it is necessary to find an expression for W, the approximating polynomial will be truncated when terms become insignificant. It is therefore necessary to test the approximation of the previous section and to find the limitations of its applicability.

The approximation is investigated by comparing the finite raytrace computed wavefront aberration curves to the curves derived from the higher order aberration coefficients derived in §5.2. The error in the approximation can then be found by subtracting the finite wavefront aberrations from the values predicted from the fifth order polynomial and the standard deviation of the approximation can be evaluated for 20 data points used to define the finite wavefront at each field position and azimuth.

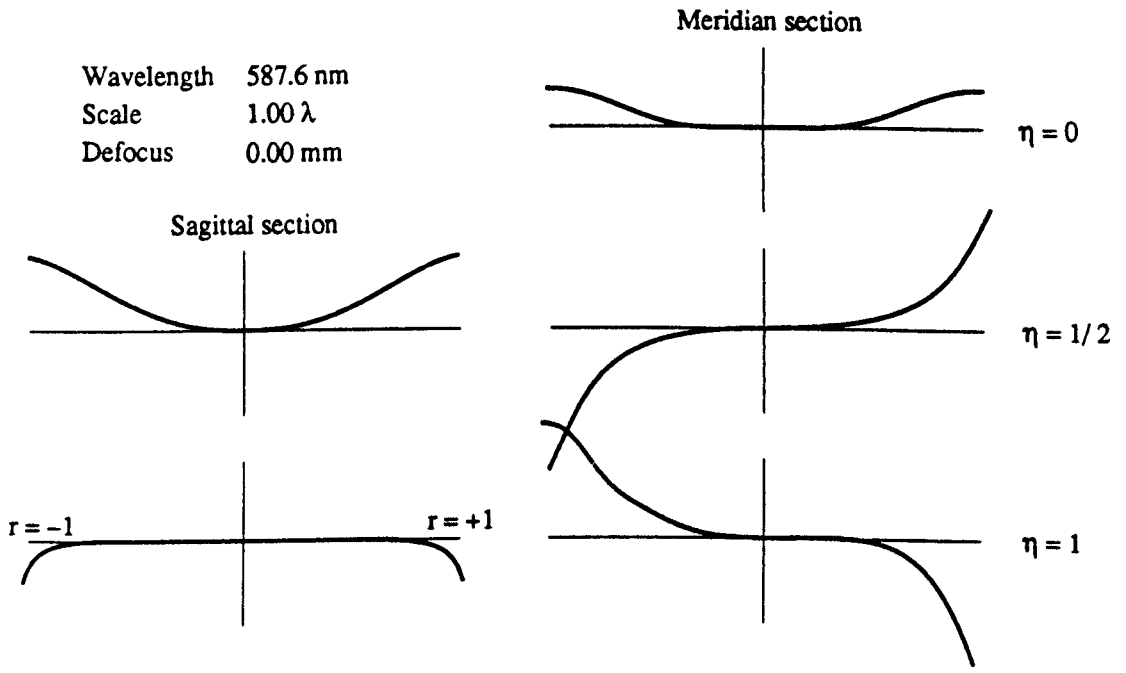


Figure 5.5. Raytrace computed wavefront aberrations for an f/4 triplet (see text).

The first example is the triplet from chapter 3, §3.4.3. The aberrations have been reduced to several wavelengths but it is not especially well corrected. It also has a particular problem, which is that coma changes sign between the zone and margin of the field (fig. 5.5).

The error in the approximation is shown graphically in fig. 5.6 and the standard deviations of the wavefronts are given in table 5.3.

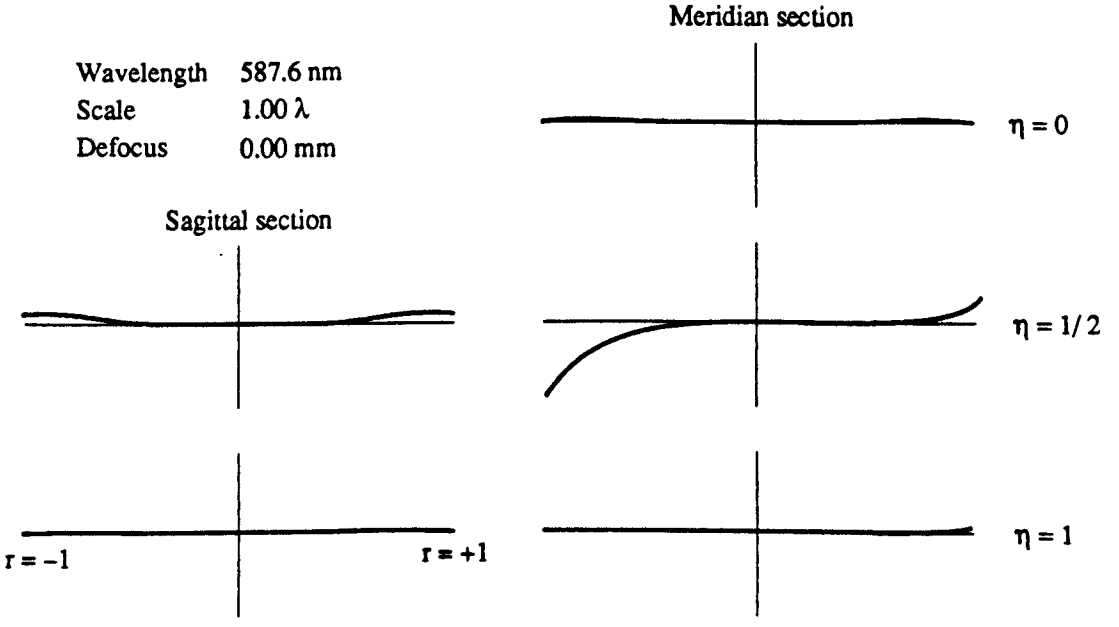


Figure 5.6. Wavefront error curves for an f/4 triplet (see text) using finite ray trace results from the margin of the field.

ϱ	ϕ	σ
0.0	0.00	0.0025
0.7	0.00	0.1088
0.7	$\pi/2$	0.0669
1.0	0.00	0.0238
1.0	$\pi/2$	0.0069

Table 5.3 Standard deviations of the approximating wavefronts for the f/4 triplet test lens.

The fifth order polynomial approximation is particularly good for the axial and the full semi-field position ($\varrho = 0$ and $\varrho = 1$) which have been used to determine the fifth order coefficients. However, at the zone of the field, the standard deviation rises to $\approx \lambda/10$. If a better approximation at the zone of the field is required, equations (5.17) to (5.33) can be recast for $\varrho = 1/\sqrt{2}$. Equations (5.17) and (5.18) become:

$$W(1,\pi/2,1/\sqrt{2}) = {}_0w_{20} + {}_0w_{40} + {}_0w_{60} + {}_0w_{80} + \tfrac{1}{2}\,{}_2w_{40} + \tfrac{1}{2}\,{}_2w_{20} + \tfrac{1}{4}\,{}_4w_{20} \tag{5.43}$$

$$W(1/\sqrt{2},\pi/2,1/\sqrt{2}) = \tfrac{1}{2}\,{}_0w_{20} + \tfrac{1}{4}\,{}_0w_{40} + \tfrac{1}{8}\,{}_0w_{60} + \tfrac{1}{16}\,{}_0w_{80} + \tfrac{1}{8}\,{}_2w_{40} + \tfrac{1}{4}\,{}_2w_{20} + \tfrac{1}{8}\,{}_4w_{20} \tag{5.44}$$

which can be solved for ${}_2w_{40}$ and ${}_4w_{20}$ to give that

$${}_2w_{40} = 4W(1,\pi/2,1/\sqrt{2}) - 8W(1/\sqrt{2},\pi/2,1/\sqrt{2}) - 2\,{}_0w_{40} - 3\,{}_0w_{60} - \tfrac{7}{2}\,{}_0w_{80} \tag{5.45}$$

and

$${}_4w_{20} = 16W(1/\sqrt{2},\pi/2,1/\sqrt{2}) - 4W(1,\pi/2,1/\sqrt{2}) - 4\,{}_0w_{20} + 2\,{}_0w_{60} + 3\,{}_0w_{80} - 2\,{}_2w_{20}. \tag{5.46}$$

Similarly, omitting details, equations (5.26) – (5.29) become

$${}_2w_{42} + \tfrac{1}{2}\,{}_4w_{22} = W(1,0,1/\sqrt{2}) + W(-1,0,1/\sqrt{2}) - 2({}_0w_{20} + {}_0w_{40} + {}_0w_{60} + {}_0w_{80}) - {}_2w_{40} - {}_2w_{20} - {}_2w_{22} - \tfrac{1}{2}\,{}_4w_{20} \tag{5.47}$$

$$2\,{}_1w_{51} + \gamma = \sqrt{2}(W(1,0,1/\sqrt{2}) - W(-1,0,1/\sqrt{2})) - 2\,{}_1w_{31} \tag{5.48}$$

$${}_2w_{42} + {}_4w_{22} = 4(W(1/\sqrt{2},0,1/\sqrt{2}) + W(-1/\sqrt{2},0,1/\sqrt{2})) - 4\,{}_0w_{20} - 2\,{}_0w_{40} - {}_0w_{60} - \tfrac{1}{2}\,{}_0w_{80} - {}_2w_{40} - 2\,{}_2w_{20} - 2\,{}_2w_{22} - {}_4w_{20} \tag{5.49}$$

$${}_1w_{51} + \gamma = 4(W(1/\sqrt{2},0,1/\sqrt{2}) - W(-1/\sqrt{2},0,1/\sqrt{2})) - 2\,{}_1w_{31} \tag{5.50}$$

and by replacing the RHS's of equations (5.47 – 5.50) with the constants A', B', C' and D' the unknown aberration coefficients can be found;

$${}_2w_{42} = 2A' - C' \tag{5.51}$$

$${}_4w_{22} = 2\,(C' - A') \tag{5.52}$$

$${}_1w_{51} = B' - D' \tag{5.53}$$

$$\gamma = 2D' - B'. \tag{5.54}$$

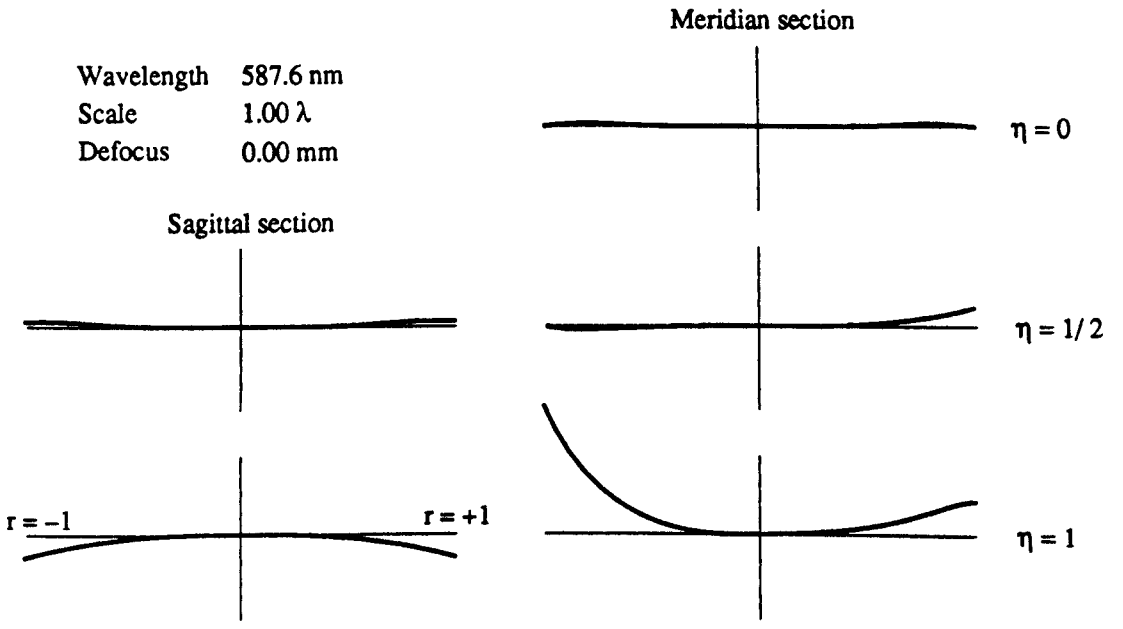


Figure 5.7. Wavefront error curves for an f/4 triplet (see text) using finite ray trace results for the zone of the field.

If these equations are used to determine the fifth order aberration coefficients, then it would be reasonable to expect a better approximation at the zone of the field. Figure 5.7 shows the error in the approximation for the triplet lens using the zone of the field for the fitting points.

This approximation is not as good as that shown in fig. 5.6 and does not even have zero values at the zone and margin of the aperture, which would be expected and was obtained for the approximation shown in fig. 5.6. A possible reason for this is that in equations (5.45) and (5.46), comparatively large numbers are being subtracted to obtain a value for the aberration coefficients which may well be less than a wavelength in magnitude. Any errors at this stage will produce errors in the coefficients derived for the meridian section. Fig. 5.7 demonstrates that the sagittal curves have errors that are greater than their counterparts in fig. 5.6. The errors in the meridian section are even greater still and demonstrate a possible propagation of error from the results obtained from the sagittal section.

This technique has also been tested on two examples which were both well corrected across 80% of the aperture. The large edge of aperture aberrations ($\approx 100\lambda$) caused

the approximation to break down since four of the fitting points are aberration values at the aperture margin. An attempt was made to counter this problem by again recasting the equations to only use the finite wavefront aberration values calculated at the zone of the aperture for both off-axis field positions. Although this improved the approximation, the results were still unsatisfactory as demonstrated by the error curves in fig. 5.8.

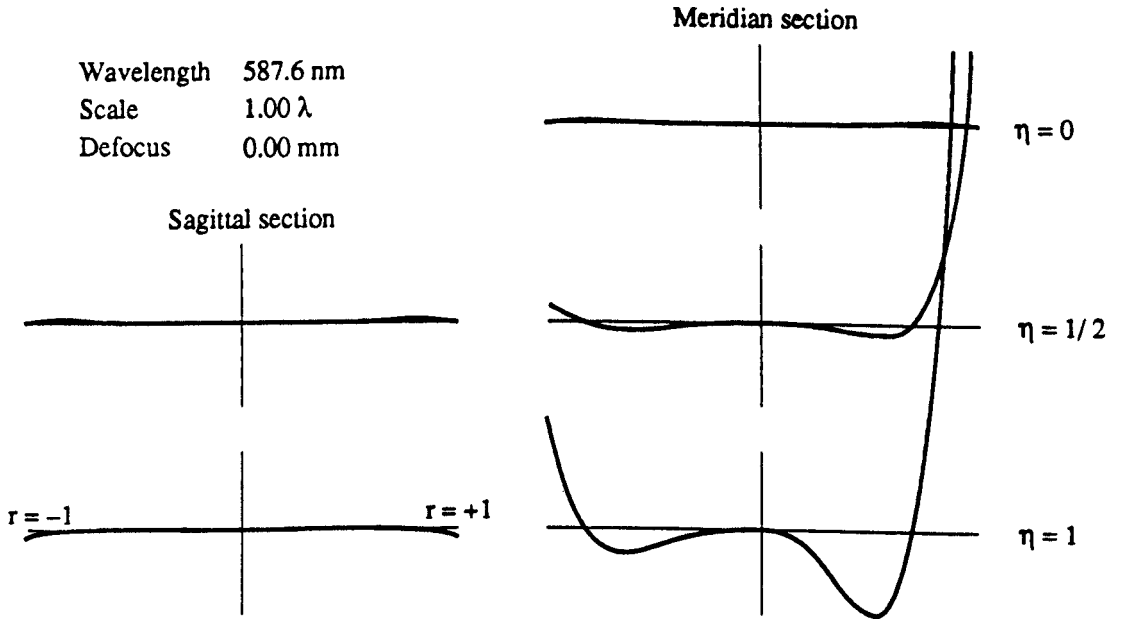


Figure 5.8. Wavefront error curves for a Tessar lens with large edge of aperture aberrations.

As a result, the approximate aberration coefficients calculated in §5.2 only form a good approximation to the actual wavefront when an optical system has been vignetted and the final adjustments are being made to optimize the aberration balance. The approximation breaks down when the system is badly behaved, i.e. when there are large aberration values at one or two points in the aperture or the field. Results also demonstrate that the equations derived from the full semi-field position are better conditioned and introduce less error than the coefficients derived

from zone of the field. It should be noted that it is generally necessary to use finite raytrace computed wavefront aberrations from both the zone and the margin of the aperture so that any turning points in the wavefront are well approximated.

5.4 Balancing the Wavefront to Maximize Strehl Ratio

The Strehl ratio, I, which is defined as the ratio of the maximum intensity in the aberrated point spread function to the maximum intensity in the unaberrated Airy pattern, can be written as [5.19]

$$I = \frac{1}{A}\left|\int_A e^{ikW} dA\right|^2 \tag{5.55}$$

where W is the wavefront aberration and it is assumed that the integration is performed over the area of a circularly symmetric exit pupil. Maréchal [5.20] bas shown that equation (5.55) can be approximated by

$$I \sim \left|1 - \frac{2\pi^2}{\lambda^2} E\right|^2 . \tag{5.56}$$

for values of the Strehl ratio, $I \geq 0.8$, where E is the variance of the wavefront aberration. Therefore, minimizing the variance of the wavefront aberration will maximize the Strehl ratio. King [5.10] subsequently demonstrated that minimizing the variance of the wavefront aberration is a useful measure of image quality for systems with $I \geq 0.5$, thus greatly extending the range of applicability of the variance as a meaningful measure of image quality.

The variance, E, is defined as the difference between the mean square value of the wavefront aberration and the mean value of the wavefront aberration squared,

$$E = \frac{1}{A}\int_A \int W^2 dA - \left[\frac{1}{A}\int_A \int W dA\right]^2 \tag{5.57}$$

which, when expressed in polar coordinates becomes

$$E = \frac{1}{\pi^2} \int_0^{2\pi} \int_0^1 W^2 r \, dr \, d\phi - \left[\frac{1}{\pi^2} \int_0^{2\pi} \int_0^1 W r \, dr \, d\phi \right]^2 . \tag{5.58}$$

The variance of the total wavefront aberration can be expressed as the sum of the variances of the even and odd parts of the wavefront aberration. This has suggested the approach followed by Hopkins [5.6], King [5.10] and others, where expressions are derived for the optimum ratio of the third to fifth order aberration coefficients that maximize the Strehl ratio.

In the analyses of Hopkins [5.6] and King [5.10], the even contribution to the wavefront aberration has been considered to comprise the third and fifth order spherical aberration terms together with defocus, and the odd component, the third and fifth order coma terms and the transverse focal shift, ${}_1w_{11}$. However, it is also possible to consider balancing the astigmatism and field curvature terms and to also include oblique spherical aberration, elliptical coma and distortion in the analysis. This is possible since the results of §5.2 allow all fifth order terms to be approximated and also Appendix G provides a proof that the total variance of the wavefront aberration is equal to the sum of the variances of *several* odd and even components.

The wavefront aberration considered in this analysis consists of two even components, spherical aberration and astigmatism and field curvature, and two odd components, coma and distortion. In addition, the longitudinal defocus term is included when optimizing the spherical aberration balance. However, the transverse focal shift term ${}_1w_{11}$ is neglected, since in practice it is not a degree of freedom that can be used when viewing extended objects in practical optical systems.

The even part of $W(r, \phi, \varrho)$, which contributes spherical aberration terms on axis, is given by

$$W(r, \phi, \varrho) = {}_0w_{20} r^2 + {}_0w_{40} r^4 + {}_0w_{60} r^6 + {}_0w_{80} r^8 \tag{5.59}$$

where ${}_0w_{20}$ has been included so that the position of the best focal plane may also be calculated. If equation (5.59) is substituted into equation (5.58) and the integrals evaluated, an expression is obtained for the variance of this wavefront;

$$\begin{aligned} E = {} & \tfrac{1}{12}\,{}_0w_{20}^2 + \tfrac{4}{45}\,{}_0w_{40}^2 + \tfrac{9}{112}\,{}_0w_{60}^2 + \tfrac{16}{225}\,{}_0w_{80}^2 \\ & + \tfrac{1}{6}\,{}_0w_{20}\,{}_0w_{40} + \tfrac{3}{20}\,{}_0w_{20}\,{}_0w_{60} + \tfrac{2}{15}\,{}_0w_{20}\,{}_0w_{80} \\ & + \tfrac{1}{6}\,{}_0w_{40}\,{}_0w_{60} + \tfrac{16}{105}\,{}_0w_{60}\,{}_0w_{80} + \tfrac{3}{20}\,{}_0w_{60}\,{}_0w_{80}. \end{aligned} \tag{5.60}$$

Assuming that the changes required to optimize the Strehl ratio are small and that the higher order aberrations remain effectively static during these changes, the variance can be minimized by changing the third order spherical aberration coefficient ${}_0w_{40}$ and the position of the focal plane ${}_0w_{20}$. This assumption is not unreasonable since it is known that the higher order aberrations vary more slowly than the third order aberrations. The partial differentials of the variance with respect to ${}_0w_{20}$ and ${}_0w_{40}$ are given by

$$\frac{\partial E}{\partial\,{}_0w_{20}} = \tfrac{1}{6}\,{}_0w_{20} + \tfrac{1}{6}\,{}_0w_{40} + \tfrac{3}{20}\,{}_0w_{60} + \tfrac{2}{15}\,{}_0w_{80} \tag{5.61}$$

$$\frac{\partial E}{\partial\,{}_0w_{40}} = \tfrac{1}{6}\,{}_0w_{20} + \tfrac{8}{45}\,{}_0w_{40} + \tfrac{1}{6}\,{}_0w_{60} + \tfrac{16}{105}\,{}_0w_{80} \tag{5.62}$$

and the variance has stationary values when equations (5.61) and (5.62) are zero. Solving for the value of ${}_0w_{20}$ and ${}_0w_{40}$ gives

$${}_0w_{20} = \tfrac{3}{5}\,{}_0w_{60} + \tfrac{96}{105}\,{}_0w_{80} \tag{5.63}$$

and

$${}_0w_{40} = -\left\{\tfrac{3}{2}\,{}_0w_{40} + \tfrac{180}{105}\,{}_0w_{80}\right\}. \tag{5.64}$$

These values minimize the variance since the second order partial differentials, which are given by

$$\frac{\partial^2 E}{\partial\,{}_0w_{40}^2} = \frac{8}{45} \tag{5.65}$$

$$\frac{\partial^2 E}{\partial\,{}_0w_{20}^2} = \frac{1}{6} \tag{5.66}$$

$$\frac{\partial^2 E}{\partial\,{}_0w_{40}\partial\,{}_0w_{20}} = \frac{1}{6} \tag{5.67}$$

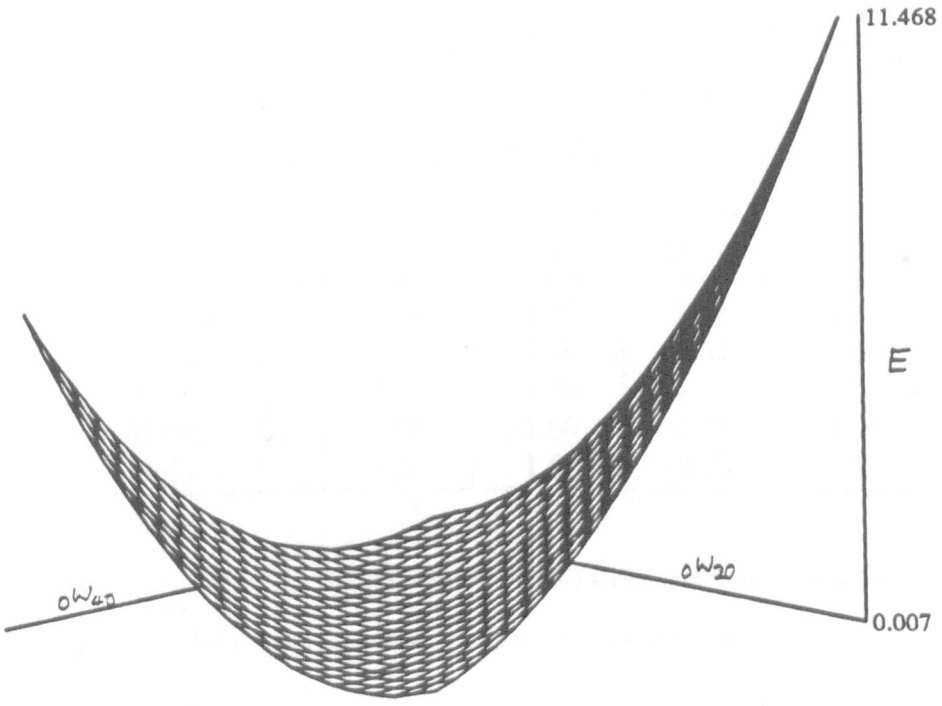


Figure 5.9 Surface profile map of the change in variance, E, as $_0w_{20}$ and $_0w_{40}$ are varied.

satisfy the two conditions,

$$\frac{\partial^2 E}{\partial\, {}_0w_{40}^2}\frac{\partial^2 E}{\partial\, {}_0w_{20}^2} - \left(\frac{\partial^2 E}{\partial\, {}_0w_{40}\partial\, {}_0w_{20}}\right)^2 > 0 \qquad (5.68)$$

$$\frac{\partial^2 E}{\partial\, {}_0w_{40}^2} > 0 \qquad (5.69)$$

which are the necessary and sufficient conditions for a function of two variables to have a minimum [5.21].

Figure 5.9 shows a profile plot of the variance as $_0w_{20}$ and $_0w_{40}$ are varied between -5.0 and $+5.0$ wavelengths. The values of $_0w_{60}$ and $_0w_{80}$ were both fixed at $+1.0$ wavelength. The plot is interesting since it demonstrates that a valley exists along which the variance increases slowly. This implies that a series of values exist for $_0w_{20}$ and $_0w_{40}$ for which the variance remains close to the analytical minimum.

The existence of this valley can be examined further by computing the values of

the variance at all the mesh points and filtering the data to plot only those values below a specified level. Figure 5.10 shows a scatter plot of the thresholded variance values for five different combinations of values for ${}_0w_{60}$ and ${}_0w_{80}$ which are given in table 5.4. The values of ${}_0w_{20}$ and ${}_0w_{40}$ give the analytical solution which has been calculated from equations (5.63) and (5.64).

Balance	${}_0w_{60}$	${}_0w_{80}$	${}_0w_{20}$	${}_0w_{40}$
Balance A	1.00	1.00	1.51	-3.21
Balance B	1.00	2.00	2.43	-4.93
Balance C	3.00	1.00	2.71	-6.21
Balance D	2.00	-1.00	0.28	-1.29
Balance E	-3.00	1.00	-0.89	2.79

Table 5.4. Values for the higher order aberration coefficients in wavelengths for the axial aberration balances shown in figure 5.10.

All the valleys lie at 45 degrees to the axes and hence the valley floors can be determined by a linear equation passing through the analytical minimum solution and with a gradient of -1. For example, if the variance has a minimum at $({}_0w'_{20},\ {}_0w'_{40})$ then the valley floor is given by

$$ {}_0w_{40} = -\,{}_0w_{20} + {}_0w'_{20} + {}_0w'_{40} \tag{5.70}$$

and hence if an optimization program cannot find the value of ${}_0w_{40}$ which achieves the analytical minimum, then ${}_0w_{20}$ can be calculated from equation (5.70) to achieve the best possible solution.

The shape of the wavefront aberration curves when the above axial spherical aberration balance has been achieved is shown in figure 5.11. The shape of these curves remains the same as the values of ${}_0w_{60}$ and ${}_0w_{80}$ increase in magnitude, but the minimum achievable variance increases causing a drop in the Strehl ratio.

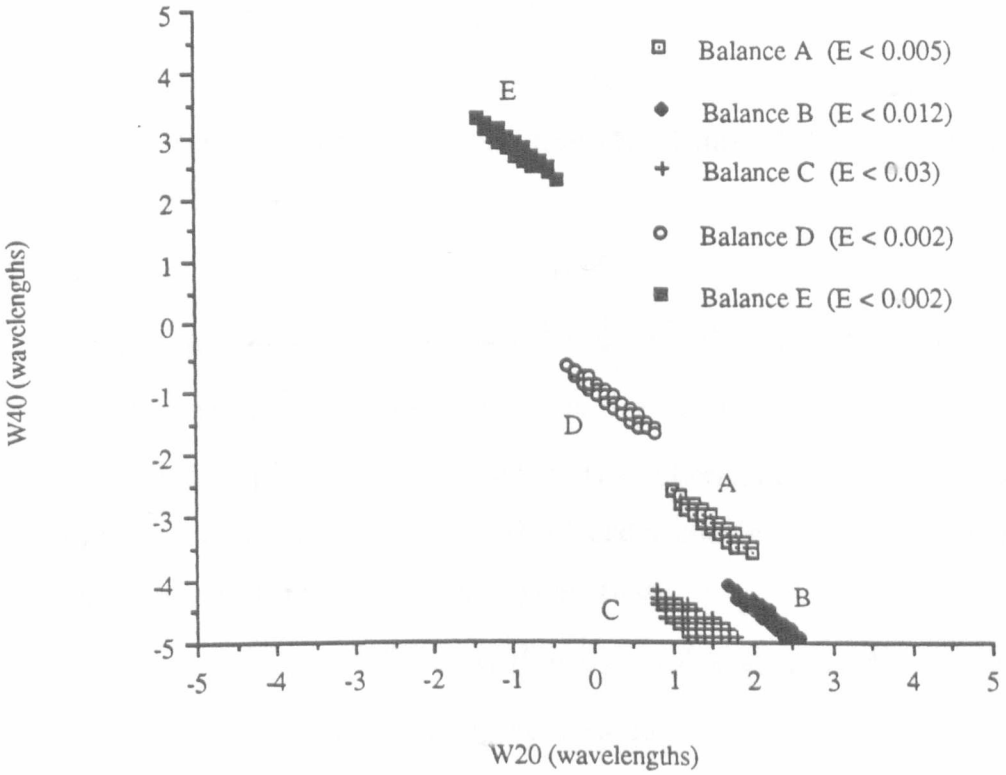


Figure 5.10 Scatter plot of filtered variance values for different axial aberration balances.

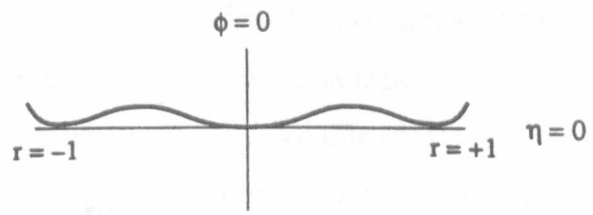


Figure 5.11. Theoretical wavefront aberrations for an axial aberration balance which maximizes the Strehl ratio.

Coma contributes odd terms to the wavefront aberration and hence W is given to fifth order by

$$W = {}_1w_{31}\varrho r^3\cos\phi + {}_1w_{51}\varrho r^5\cos\phi + {}_3w_{31}\varrho^3 r^3\cos\phi + {}_3w_{33}\varrho^3 r^3\cos^3\phi \quad (5.71).$$

where the last two terms are the elliptical coma coefficients. The variance of this expression is given by

$$E = \frac{1}{\pi^2} \int_0^{2\pi} \int_0^1 W^2 r \, dr \, d\phi. \tag{5.72}$$

since the integral of the odd functions in the second term on the right-hand side of equation (5.58) is zero. Substituting equation (5.71) into (5.72) and integrating yields an expression for the variance, E,

$$\begin{aligned} E = {} & \tfrac{1}{8}\,{}_1w_{31}^2\varrho^2 + \tfrac{1}{5}\,{}_1w_{31}\,{}_1w_{51}\varrho^2 + \tfrac{1}{12}\,{}_1w_{51}^2\varrho^2 \\ & + \tfrac{1}{4}\,{}_1w_{31}\,{}_3w_{31}\varrho^4 + \tfrac{1}{5}\,{}_1w_{51}\,{}_3w_{31}\varrho^4 + \tfrac{1}{8}\,{}_3w_{31}^2\varrho^3 + \tfrac{3}{16}\,{}_1w_{31}\,{}_3w_{33}\varrho^4 \\ & + \tfrac{3}{20}\,{}_1w_{51}\,{}_3w_{33}\varrho^4 + \tfrac{3}{16}\,{}_3w_{31}\,{}_3w_{33}\varrho^6 + \tfrac{5}{64}\,{}_3w_{33}^2\varrho^6. \end{aligned} \tag{5.73}$$

Following an identical approach to that used for the spherical aberration terms, the variance is minimized with respect to the third order coefficient ${}_1w_{31}$. Differentiating equation (5.56) with respect to ${}_1w_{31}$ gives that

$$\frac{\partial E}{\partial\, {}_1w_{31}} = \tfrac{1}{4}\,{}_1w_{31}\varrho^2 + \tfrac{1}{5}\,{}_1w_{51}\varrho^2 + \tfrac{1}{4}\,{}_3w_{31}\varrho^4 + \tfrac{3}{16}\,{}_3w_{33}\varrho^4 \tag{5.74}$$

and hence for the variance to have a stationary value

$${}_1w_{31} = -\left\{\tfrac{4}{5}\,{}_1w_{51} + \left({}_3w_{31} + \tfrac{3}{4}\,{}_3w_{33}\right)\varrho^2\right\} \tag{5.75}$$

which is a minimum value since the second order partial derivative is positive. The second term in equation (5.75) cannot be replaced by γ which was defined in §5.2 as the sum of the elliptical coma terms since the elliptical coma terms do not appear in equation (5.75) in the ratio of 1:1. There are two possible ways of handling this; First, the ratio of 4:3 may be considered close enough to the required ratio of 1:1 so as not to be significant, or second, an assumption could be made about the distribution of magnitudes between the two coefficients. The former approach has been adopted and produces acceptable results for the lenses that have been tested.

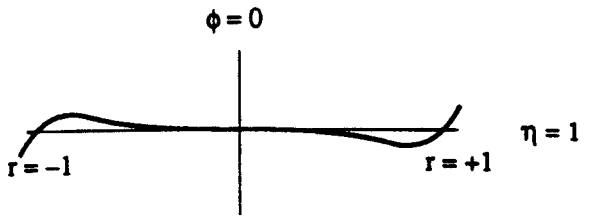


Figure 5.12. Theoretical balance of the coma terms in the wavefront aberration to maximise Strehl ratio.

The shape of the theoretical wavefront aberration curves which have been balanced for the marginal field position are given in figure 5.12. The general form of the curves is unaffected by the magnitude of the fifth order coefficients. However, the minimum achievable variance increases as the magnitude of the fifth order coefficients increase.

Oblique spherical aberration, astigmatism and field curvature all contribute even terms to the wavefront aberration at off-axis field positions. The wavefront aberration due to these terms is given by

$$W = {}_2w_{40}\varrho^2r^4 + {}_2w_{42}\varrho^2r^4\cos^2\phi + {}_2w_{20}\varrho^2r^2 + {}_2w_{22}\varrho^2r^2\cos^2\phi + {}_4w_{20}\varrho^4r^2 + {}_4w_{22}\varrho^4r^2\cos^2\phi \tag{5.76}$$

which when substituted in equation (5.58) and the variance evaluated yields

$$\begin{aligned} E = {} & \tfrac{1}{12}\,{}_2w_{20}^2\varrho^4 + \tfrac{1}{16}\,{}_2w_{22}^2\varrho^4 + \tfrac{4}{45}\,{}_2w_{40}\varrho^4 + \tfrac{17}{360}\,{}_2w_{42}\varrho^4 + \tfrac{1}{12}\,{}_4w_{20}^2\varrho^4 \\ & + \tfrac{1}{16}\,{}_4w_{22}^2\varrho^4 + \tfrac{1}{12}\,{}_2w_{20}\,{}_2w_{22}\varrho^4 + \tfrac{1}{6}\,{}_2w_{20}\,{}_2w_{40}\varrho^4 + \tfrac{1}{12}\,{}_2w_{20}\,{}_2w_{42}\varrho^4 \\ & + \tfrac{1}{6}\,{}_2w_{20}\,{}_4w_{20}\varrho^6 + \tfrac{1}{12}\,{}_2w_{20}\,{}_4w_{22}\varrho^6 + \tfrac{1}{12}\,{}_2w_{22}\,{}_2w_{40}\varrho^4 + \tfrac{5}{48}\,{}_2w_{22}\,{}_2w_{42}\varrho^4 \\ & + \tfrac{1}{12}\,{}_2w_{22}\,{}_4w_{20}\varrho^6 + \tfrac{1}{8}\,{}_2w_{22}\,{}_4w_{22}\varrho^6 + \tfrac{4}{45}\,{}_2w_{40}\,{}_2w_{42}\varrho^4 + \tfrac{1}{6}\,{}_2w_{40}\,{}_4w_{20}\varrho^6 \\ & + \tfrac{1}{12}\,{}_2w_{40}\,{}_4w_{22}\varrho^6 + \tfrac{1}{12}\,{}_2w_{42}\,{}_4w_{20}\varrho^6 + \tfrac{5}{48}\,{}_2w_{42}\,{}_4w_{22}\varrho^6 + \tfrac{1}{12}\,{}_4w_{20}\,{}_4w_{22}\varrho^8. \end{aligned} \tag{5.77}$$

Both ${}_2w_{20}$ and ${}_2w_{22}$ can be varied to minimize the variance and therefore the partial differentials of equation (5.77) are taken with respect to these two variables:

$$\begin{aligned} \frac{\partial E}{\partial\,{}_2w_{20}} = {} & \tfrac{1}{6}\,{}_2w_{20}\varrho^4 + \tfrac{1}{12}\,{}_2w_{22}\varrho^6 + \tfrac{1}{6}\,{}_2w_{40}\varrho^4 + \tfrac{1}{12}\,{}_2w_{42}\varrho^4 \\ & + \tfrac{1}{6}\,{}_4w_{20}\varrho^6 + \tfrac{1}{12}\,{}_4w_{22}\varrho^6 \end{aligned} \tag{5.78}$$

$$\begin{aligned} \frac{\partial E}{\partial\,{}_2w_{22}} = {} & \tfrac{1}{12}\,{}_2w_{20}\varrho^4 + \tfrac{1}{8}\,{}_2w_{22}\varrho^4 + \tfrac{1}{12}\,{}_2w_{40}\varrho^4 + \tfrac{5}{48}\,{}_2w_{42}\varrho^4 \\ & + \tfrac{1}{12}\,{}_4w_{20}\varrho^6 + \tfrac{1}{8}\,{}_2w_{42}\varrho^6 \end{aligned} \tag{5.79}$$

Equations (5.78) and (5.79) must be simultaneously zero for the variance to have a stationary value and therefore solving for ${}_2w_{20}$ and ${}_2w_{22}$ gives

$${}_2w_{20} = -\left\{\,{}_2w_{40} + \tfrac{1}{8}\,{}_2w_{42} + {}_4w_{20}\varrho^2\right\} \tag{5.80}$$

and

$$ {}_2w_{22} = -\left\{ \tfrac{3}{4}\, {}_2w_{42} + {}_4w_{22}\varrho^2 \right\}. \tag{5.81} $$

These values define a minimum value for the variance since the conditions in equations (5.68) and (5.69) are again satisfied.

A plot of the variance surface as ${}_2w_{20}$ and ${}_2w_{22}$ are varied between -5.0 and $+5.0$ wavelengths is shown in figure 5.13. The surface is approximately rotationally symmetric about the analytical minimum value and therefore there is an even degradation in the Strehl ratio as the values of ${}_2w_{20}$ and ${}_2w_{22}$ move away from this point.

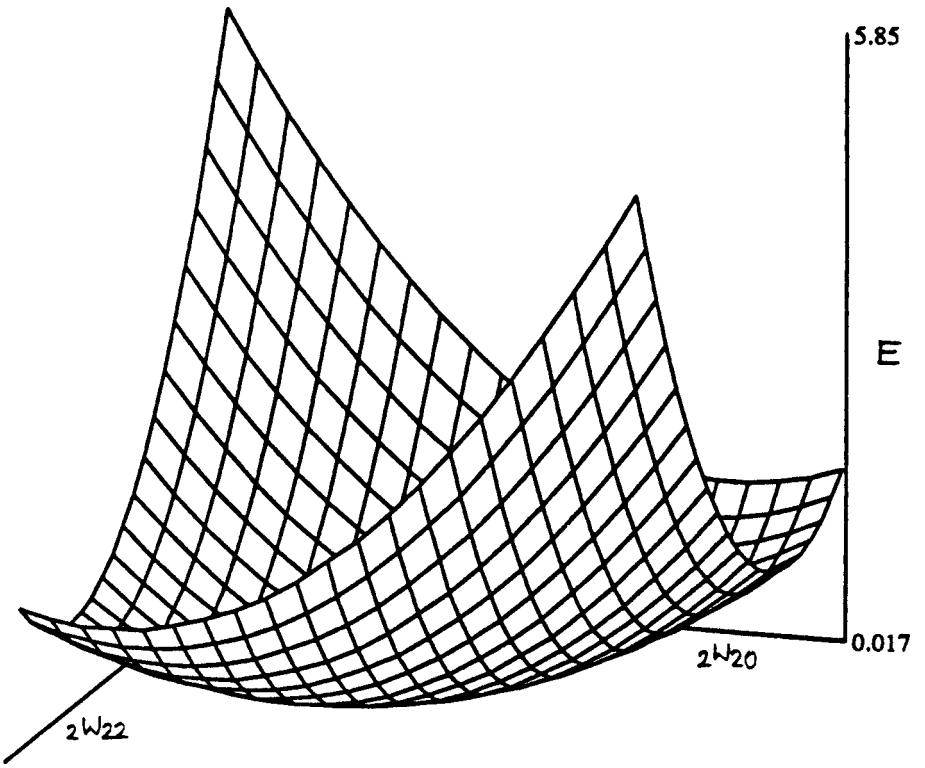


Figure 5.13. Surface profile of the change in variance, E, as ${}_2w_{20}$ and ${}_2w_{22}$ are varied.

Further confirmation of the symmetrical increase in the variance is provided by fig-

ure 5.14 which shows four scatter plots for different aberration balances detailed in table 5.5. The values of ${}_2w_{22}$ and ${}_2w_{20}$ which appear in the table have been calculated from equations (5.80) and (5.81) and therefore give the analytical solution. To obtain the scatter graph, all mesh points on the contour map have been computed and the data filtered so that only values of the variance below a specified threshold level appear. The areas are approximately circular although some elongation at 45 degrees to the axes is evident. This effect is not as marked as for the axial spherical aberration balancing and therefore cannot be exploited.

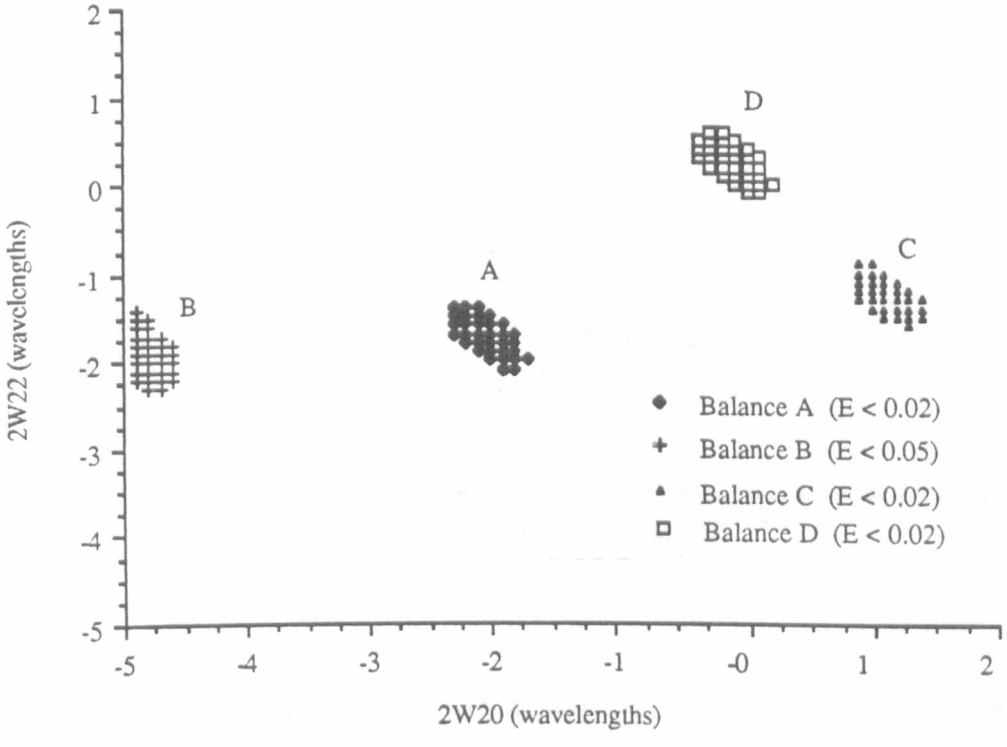


Figure 5.14 Scatter plot of filtered variance values
for different off-axis aberration balances.

The values of the fifth order coefficients and the values of the third order coefficients for the analytical minimum for the different balances shown in figure 5.14 are given

in table 5.5;

Balance	$_2w_{40}$	$_2w_{42}$	$_4w_{20}$	$_4w_{22}$	$_2w_{20}$	$_2w_{22}$
Balance A	1.00	1.00	1.00	1.00	-2.13	-1.75
Balance B	2.00	1.00	3.00	1.00	-5.13	-1.75
Balance C	2.00	-1.00	-3.00	2.00	1.13	-1.25
Balance D	1.00	1.00	-1.00	-1.00	-0.13	0.25

Table 5.5. Values for the higher order aberration coefficients in wavelengths for the off-axis aberration balances shown in figure 5.14.

The theoretical shape of the balanced wavefront aberration curves for the one specified field position, both equations (5.68) and (5.69) have a dependence on ϱ^2, are shown in figure 5.15.

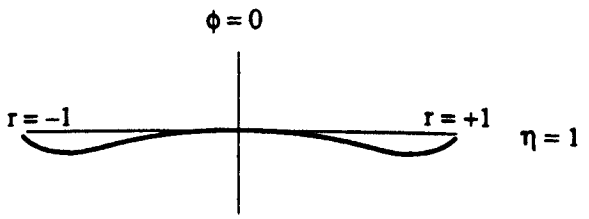


Figure 5.15. Theoretical wavefront aberrations for an off-axis aberration balance which maximizes the Strehl ratio.

Distortion contributes odd terms to the wavefront aberration and therefore equation (5.54) for variance applies. However, distortion does not affect image sharpness and is not usually toleranced on a wavefront aberration basis. The distortion balance is included here for completeness.

The wavefront aberration contribution contains the third and fifth order distortion terms and is given by

$$W = {}_3w_{11}\varrho^3 r\cos\phi + {}_5w_{11}\varrho^5 r\cos\phi \qquad (5.82)$$

which has a variance of

$$E = \tfrac{1}{4}\,{}_3w_{11}^2\varrho^6 + \tfrac{1}{2}\,{}_3w_{11}\,{}_5w_{11}\varrho^8 + \tfrac{1}{4}\,{}_5w_{11}^2\varrho^{10}. \tag{5.83}$$

The minimum variance is found when

$$\frac{\partial E}{\partial\,{}_3w_{11}} = \tfrac{1}{2}\,{}_3w_{11}\varrho^6 + \tfrac{1}{2}\,{}_5w_{11}\varrho^8 = 0 \tag{5.84}$$

Therefore

$${}_3w_{11} = -\,{}_5w_{11}\varrho^2 \tag{5.85}$$

and an exact balance can be achieved at one off-axis field position.

5.5 Examples

The correction of four test lenses has been attempted using the methods developed in §5.2 - 5.4. The lenses have been chosen in an attempt to form a representative set of examples and are given in Appendix E.

One of the assumptions underlying the optimization of the Strehl ratio which is described in this chapter, is that the higher order aberrations remain static while their third order counterparts are adjusted. In practice, this will not be the case, although it is a good approximation and a necessary one if the optimization is to converge. The small changes in the higher order aberrations are dealt with by repeating the optimization until the requested changes in the third order aberrations become negligible. This is achieved in approximately 3-4 iterations for the lenses that have been tested; a typical variation of the Seidel aberrations is shown in figure 5.16. Although the changes in the system parameters are small, it is likely that the focal length may drift by a mm or so. Focal length is not targeted since it would increase the ill-conditioning of the linear equations (§3.3.2). Scaling the focal length during optimization was tested and led to oscillation of the aberrations about their target values and so no control is placed on focal length in the present program.

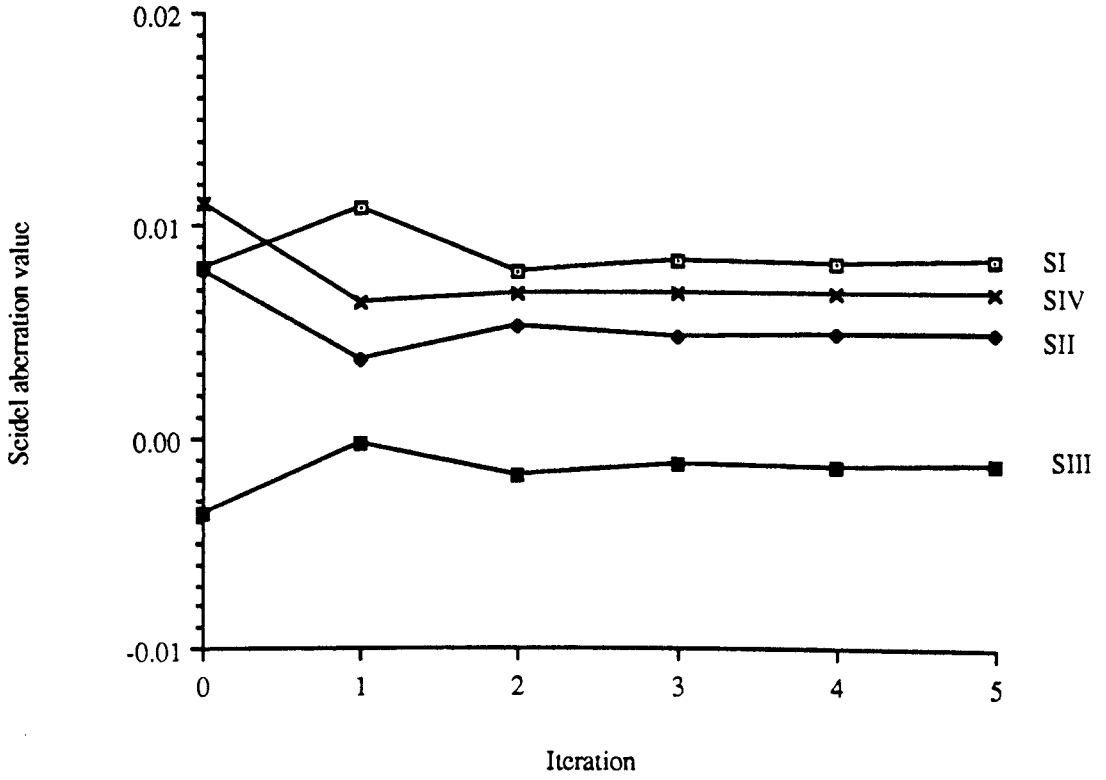


Figure 5.16. Typical variation of the Seidel aberrations during Strehl optimization.

In the examples, the higher order aberrations have been calculated using the same field position that was being optimized. This improves the accuracy of the aberration coefficients for the specified field position (§5.3) and hence the resultant spatial frequency response of the optimized system.

The first test lens was an f/4 triplet (TL5.1 in Appendix E) which has been optimized using the damped least squares (DLS) algorithm with default weighting factors. The design has a relative aperture of f/4 and field of ±15 degrees. Optimization showed that this lens has a difficult aberration balance with respect to the field coordinate (coma changes sign between the zone and margin of the field) which is a characteristic of this type of design. The wavefront aberration curves are shown in figure 5.17 and the resulting OTF (not shown) was poor with the contrast falling to below 0.2 for frequencies over $50\,c\,mm^{-1}$ at both off-axis field positions. The diffraction limited cut off frequency was $425\,c\,mm^{-1}$. Strehl optimization at the field margin produced only a small increase in the spatial frequency response for the sagittal section with the meridian section performance remaining almost unaffected. However, inspection of the wavefront aberration curves showed a reduction

from $\pm 2\lambda$ to $\pm 1\lambda$ in the meridian section (fig. 5.18). However, optimization at the zone of the field produced a significant improvement in the spatial frequency response with a slightly superior performance in the sagittal section (fig. 5.19).

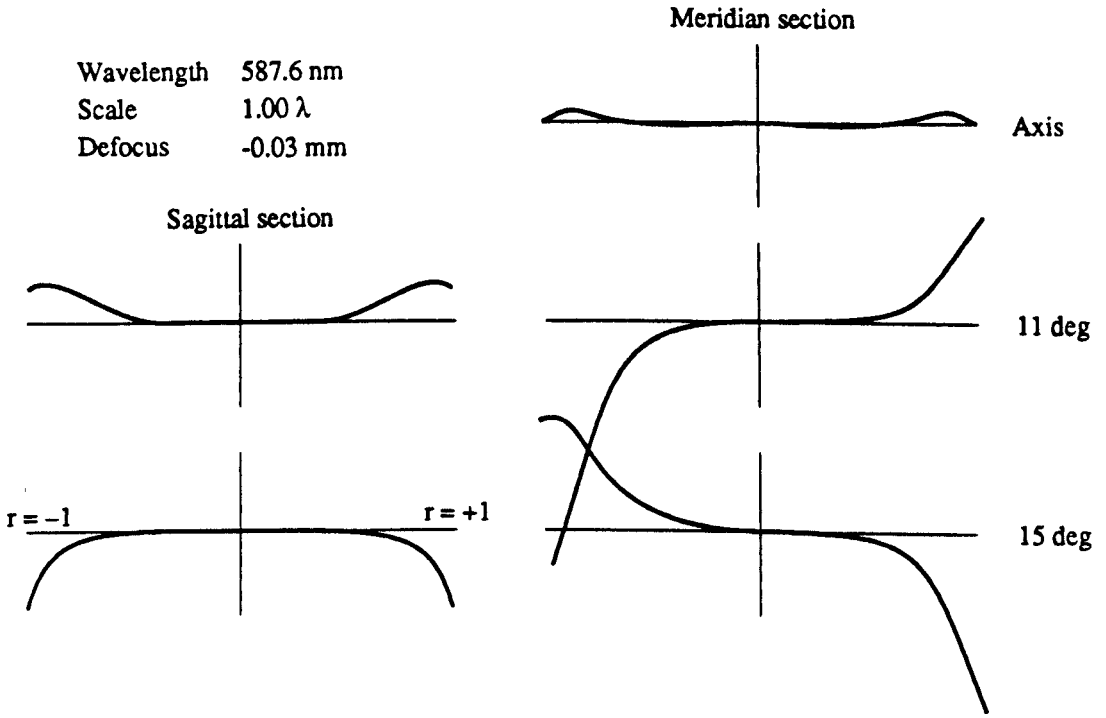


Figure 5.17. Raytrace computed wavefront aberrations for TL 5.1 (see text).

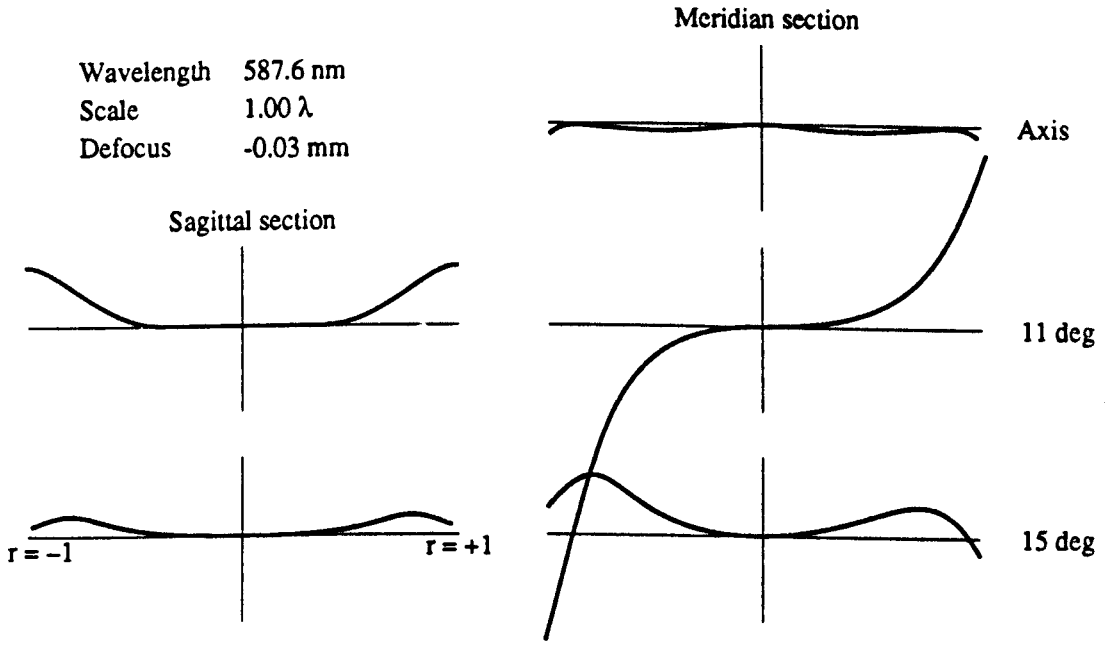
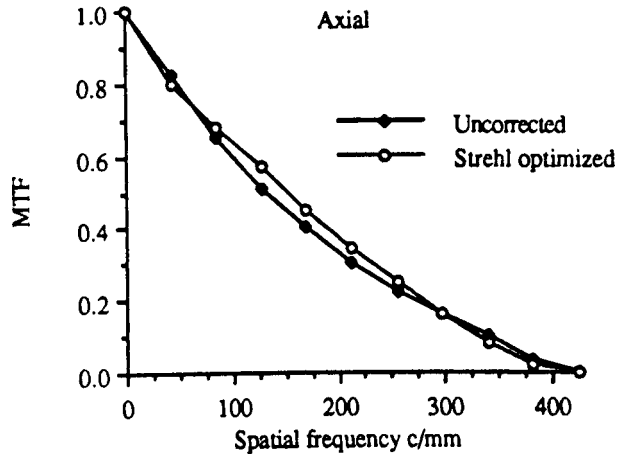
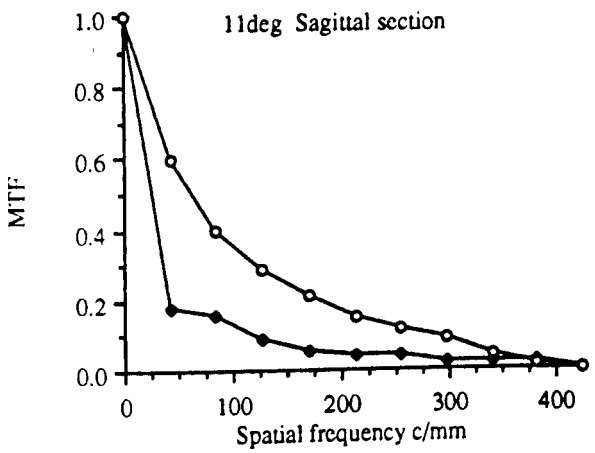
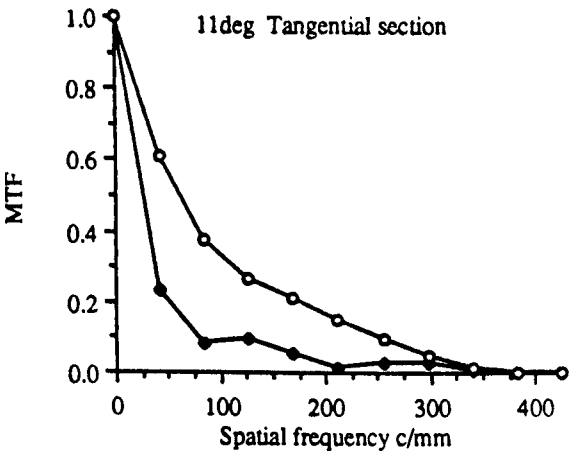
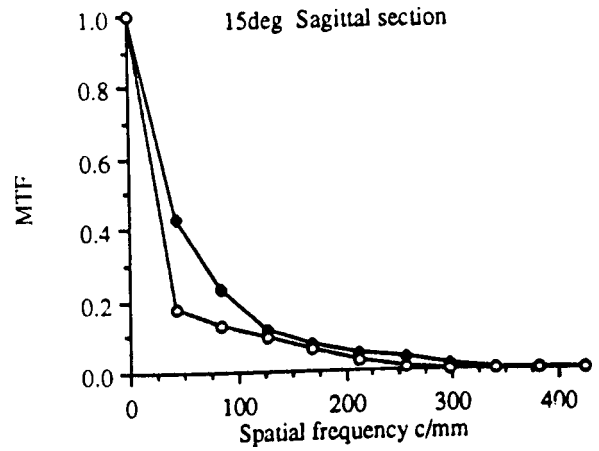


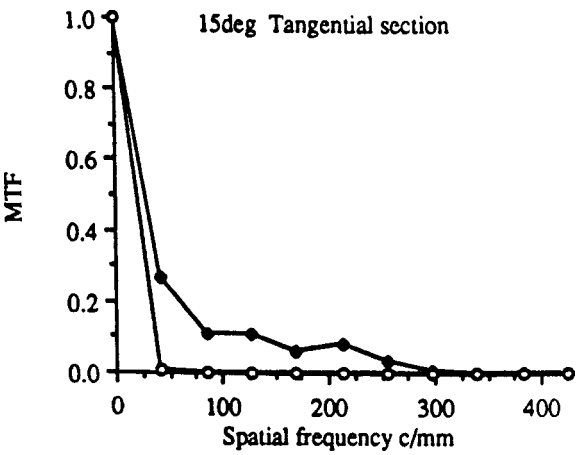
Figure 5.18. Raytrace computed wavefront aberrations for TL 5.1 after Strehl optimization.









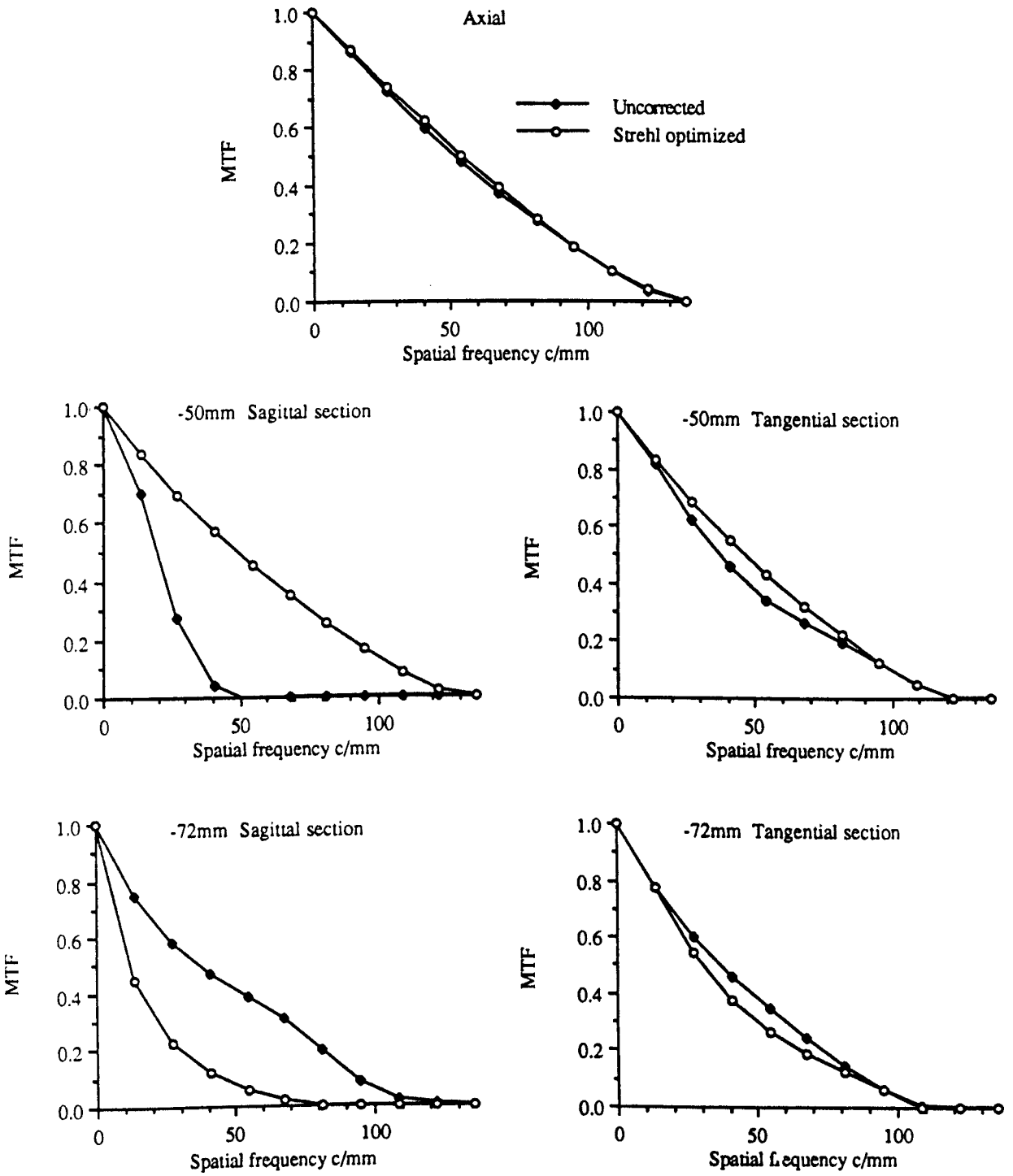


5.19. Spatial frequency response of TL5.1 both before and after Strehl optimization.

The potential problems with applying this method to the first test lens is that firstly, the lens is not close to the diffraction limit at the off-axis field positions and secondly, the approximate higher order coefficients are not as accurate as in some of the other lenses that have been tested, due to the field varying aberrations.

The second test lens was a unit magnification lens of the split Dagor type, which has been optimized using the damped least squares algorithm (lens TL5.2 in Appendix E). It is already close to the diffraction limit although the contrast drops rapidly in the sagittal section at the field zone. Strehl optimization at the zone of the field produced diffraction limited performance in both the sagittal and the meridian sections (fig. 5.20). The field margin was already close to the diffraction limit and little improvement was possible. This example clearly shows the effect of optimizing at a particular field position and illustrates the potential for developing the algorithm further by averaging the balance over the field or weighting the different field positions.

The third test lens was an f/2 Petzval lens (TL5.3 Appendix E) which has again been optimized using the damped least squares algorithm. This example has a comparatively high relative aperture but the field requirements are low (± 6 degrees). The wavefront aberration curves appear to have a reasonable balance with turning points towards the edge of the aperture. However, the magnitude of the aberrations ($\pm 4\lambda$) takes the lens well away from the diffraction limit (cut off frequency $851\, c\,mm^{-1}$) with the spatial frequency response falling to zero at $80\, c\,mm^{-1}$ in the sagittal section. Optimization at both off-axis field positions produced an improvement in the sagittal section but at the expense of performance in the meridian section. The axial spatial frequency response, which has been corrected to the diffraction limit in the three previous examples, was only slightly improved. The higher order aberrations in this lens are still too high for a diffraction limited response to be attainable. It also demonstrates that the technique is limited to $I \geq 0.5$ since the higher order aberrations remain effectively static during correction and are not reduced further.



5.20. Spatial frequency response of TL5.2 both before and after Strehl optimization.

The final example was a $25mm$ focal length Erfle eyepiece (TL5.4 Appendix E) whose performance was approaching the diffraction limit (cut off frequency $340\,c\,mm^{-1}$). Optimization of this lens was unsuccessful due to the very large requested changes in the Seidel aberrations. The lens is also particularly sensitive to changes in the curvatures of the eye lens, some of which are working close to total internal refraction. The very large changes in the Seidel aberrations are still not properly understood.

5.6 Discussion

§5.5 has shown that the optimization of Strehl ratio by balancing the third order aberrations against their higher order counterparts is a practical technique and can have a significant effect on some designs. Barakat and Houston [5.22] have noted that some third order aberrations can become intractable during final correction of an optical system and this can cause problems unless the intractable aberrations are recognized and removed. The indication and removal of intractable aberrations, which is a feature of the quadratic correction algorithm described in chapter 3, is one advantage of the method. In particular, it is worth recalling that the total variance is simply the sum of the variances from balancing the odd and the even terms in the wavefront aberration. Therefore, if the lens can only be partially balanced, there should be some improvement in its performance. Another advantage is the explicit control of the field balancing. The problem of balancing the aberrations with respect to field is an area for further investigation and beyond the scope of this thesis. However, some simple extensions of the methods in this chapter can be made; for example, averaging the performance over the field or weighting the different field positions. It is the author's belief that more sophisticated methods are probably necessary involving through focus OTF data. However, the limitations of the technique are that it can only successfully be employed on designs approaching the diffraction limit and that the higher order aberrations are not further reduced. Meiron [5.7] has described the application of the variance of the wavefront aberration as a merit function in damped least squares optimization. Such a technique has

the potential to reduce the third and higher order aberrations simultaneously. However, the example used to illustrate its application was a high aperture catadioptric system with only a modest field requirement (±2.25 degrees). In the author's experience, it is the oblique aberrations which are the most difficult to correct and balance and the spherical aberration terms do not usually present great problems. The example quoted may therefore be a 'best case' example since with the small field of view the optimization reduces to balancing the spherical aberration terms and defocus.

5.7 References

[5.1] A.E. Conrady, *Applied Optics and Optical Design*, Dover, New York (1960), p585.

[5.2] Rayces, J.L. "Ten Years of Design with Glatzel's Adaptive Method," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, pp75–84 (1980)

[5.3] Grey, D.S., "Orthogonal Polynomials as Lens Aberration Coefficients," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, 85–90 (1980).

[5.4] C.S. Williams and O.A. Becklund, *Introduction to the Optical Transfer Function*, John Wiley & Sons, New York (1989), p131

[5.5] W.B. King, "Use of the Modulation-Transfer Function (MTF) as an Aberration-Balancing Merit Function in Automatic Lens Design," J.O.S.A., **59**, 1155–1158 (1969).

[5.6] H.H. Hopkins, "The use of diffraction-based criteria of image quality in automatic optical design," Opt. Act., **13**, 343–369 (1966).

[5.7] J. Meiron, "The use of Merit Functions based on Wavefront Aberrations in Automatic Lens Design," Appl. Opt., **7**, 667–672 (1968).

[5.8] P.N. Robb, "Lens Design using Optical Aberration Coefficients," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, 109–118 (1980).

[5.9] B. Brixner, "Lens Design Merit Functions: rms image spot size and rms optical path difference," Appl. Opt., **17**, 715–716 (1978).

[5.10] W.B. King, "Dependence of the Strehl Ratio on the Magnitude of the Variance of the Wave Aberration," J.O.S.A., **58**, pp655–661 (1968).

[5.11] W.B. King, "Correlation between the Relative Modulation Function and the Magnitude of the Variance of the Wave-Aberration Difference Function,"J.O.S.A., **59**, 692–697 (1969).

[5.12] R.W. Gostick, "OTF-based Optimization Criteria for Automatic Optical Design",Opt. & Quant. Elec., **8**, 31–37 (1976).

[5.13] M.J. Kidger and P. Benham, "Optimization of the Optical Transfer Function," in Proc. SPIE Vol. **1354**, *International Lens Design Conference* , pp 92–96 (1990).

[5.14] B. Tatian, "Aberration Balancing Criterion for non diffraction-limited lenses," in Proc. SPIE Vol. **237**, *International Lens Design Conference* , pp 92–98 (1980).

[5.15] C.J. Woodruff, "A Comparison using Orthogonal Coefficients, of two forms of Aberration Balancing," Opt. Act., **22**, 933–941 (1975).

[5.16] H.A. Buchdahl, *Optical Aberration Coefficients*, Oxford University Press, London (1964).

[5.17] M.J. Kidger, "The Calculation of the Optical Transfer Function," Opt. Act., **25**, 665–680 (1978).

[5.18] Welford, W.T. *Aberrations of Optical Systems,* Adam Hilger Ltd. (1986), p139.

[5.19] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, 1966, p 461.

[5.20] A. Maréchal, "Études des effets combines de la diffraction et des aberrations géométriques sur l'image d'un point lumineux," Rev. d'Opt., **27**, 257–277 (1947).

[5.21] J. Dunning-Davies, *Mathematical Methods for Mathematicians, Physical Scientists and Engineers,* Ellis-Horwood (1982), pp 204–209.

[5.22] R. Barakat and A. Houston, "Transfer Function of an Optical System in the Presence of Off-Axis Aberrations," J.O.S.A., **55**, pp1142–1148 (1965).

CHAPTER 6

Summary and Conclusions

In this thesis the correction of lens systems has been attempted by targeting individual aberrations as opposed to using methods which require a single-value merit function. A new algorithm has been developed for this purpose, which utilises a second order approximation to the aberration variation and aberration control techniques based on Glatzel's adaptive method. As a result it has been called the Quadratic Adaptive Correction (QAC) algorithm. Adaptive aberration control can alter the aberrations that participate at each stage of the correction process by recognizing sources of ill-conditioning, intractable aberrations and aberrations that have drifted with respect to a specified tolerance band. The second order approximation allows significantly larger parameter changes to be made when compared to algorithms based on a linear approximation and also finds the aberration target values, where possible, with high precision. If this approach is compared with an algorithm using a merit function defined as the sum of the squares of the aberrations, several important features emerge. First, the adaptive aberration control means that the merit function surface changes as aberrations are removed or added to the matrix at each iteration. This lends a topological nature to the merit function surface and therefore a long winding valley in the merit function surface caused by ill-conditioning at one iteration may not be present at the next if the source of ill-conditioning has been removed. Second, if a merit function is not used during correction, (even though a merit function may still be defined for illustrative purposes), then it is possible to find different solutions when compared with merit function based gradient methods. Examples in chapter 3 showed that this was the case and that during certain correction runs a merit function can increase as well as

decrease, although the general tendency is for it to decrease as the system is brought under better control. If a merit function based method had followed the same set of directions in parameter space it would have become trapped in the local minimum solution.

The algorithm of chapter 3 is of general applicability and was used to correct two test lenses; a 50mm f/4 triplet with a field of ± 15 degrees and a 50mm f/1.7 double Gauss derivative with a field of ± 24 degrees. The results with the triplet compared favourably with known optimum solutions but the double Gauss derivative required a significant number of iterations for correction (34). In fact the double Gauss is regarded as an easier lens to design because it has more than enough degrees of freedom to correct the aberrations, whereas the triplet has only just enough useful variables for the correction and even then there are usually problems balancing the aberrations with respect to the field. This reduction in efficiency when dealing with larger systems has also been noted by Day [6.1]. Possible explanations are that increased ill-conditioning adversely affects the method or that requiring the aberrations to simultaneously reach an intermediate target level before reducing this level further towards the required values, provides too stringent a control.

One important feature of the QAC algorithm is that it allows the designer to investigate the correction of different aberrations and 'learn' about the design. This is very important and, in the absence of a true expert systems approach to lens design, can provide a useful design aid when more information is required about a particular system. It is something that quickly becomes apparent when using the program although it is not possible to quantify such information. The feedback provided is more direct than when using standard gradient methods because the individually targeted aberrations do not lose their identity. The further development of this technique as a diagnostic tool is worth further investigation, although a series of tests must be devised which are general enough in their applicability to make this a practical proposition.

With the exception of Glatzel's adaptive method, comparisons with other algorithms have been avoided because they are fundamentally different in their approach to the optimization problem. In addition, some versions of the DLS method have been developed over the past 35 years and represent the result of research by many workers. It would therefore be somewhat harsh to expect a new algorithm to perform well against such a well refined optimization method. Many modifications have come to light (no pun intended!) which would enhance the method further, for example, the use of different aberration coefficients (Zernike polynomials) and composite variables, both of which would have the effect of altering the design space. Sequential correction of aberrations was also used for some of the aberration corrections attempted on the test lenses in chapter 3. It has also been used in chapter 4 for the correction of mixed higher order aberrations and was found to allow the targeted aberrations to be corrected when simultaneous correction of all targeted aberrations had failed to make any improvement. The approach used for selecting the aberrations to be targeted based on the average rate of change could be made semi-automatic and warrants further investigation.

The application of the QAC algorithm to the examples that have so far been tested indicates that it may well be best suited as a design tool rather than as an optimization algorithm. In this context, optimization is defined as the reduction of aberration residuals whereas design is concerned with aberration balancing. These definitions therefore raise the question as to what many commercially available 'optimization' programs are attempting to be; it is the author's opinion that many programs at their outset were content to simply reduce aberration residuals since this was a major achievement at the time but that later developments were concerned more with the lens design process. The resulting algorithms not only reduce aberration residuals but also attempt a basic level of lens design. However, an "automatic" lens design program still appears to be only a remote possibility and with optical systems becoming increasingly diverse, it seems less possible that any such program will be able to deal with all cases.

One of the drawbacks with correcting individual aberrations is the selection of suitable target values. It is possible to make this process automatic for certain situations and it was decided to tackle two areas of design where current optimization methods are not always appropriate. These are the production of starting designs and the adjustment of the final aberration balance. Chapter 4 examined the possibility of using the higher order aberrations as a guide to finding starting solutions for a standard gradient optimization method (in this case the damped least squares algorithm). The correction of the higher order aberrations which generally have a lower rate of change with respect to the variable parameters than the third order aberrations, has been noted by several authors including Wynne [6.2] and Glatzel [6.3] as an important part of the design process. Such an approach is justified in that the higher order aberrations require the largest parameter changes for correction. Any subsequent changes needed to control the third order aberrations will be comparatively small and therefore will not significantly upset any previously achieved correction. Reduction of the higher order aberrations is important if the finished design is to have a good spatial frequency response. In chapter 5 it has been shown that the magnitude of the higher order aberrations fundamentally limits the minimum achievable variance of the wavefront aberration and hence the maximum possible Strehl ratio.

Seven mixed higher order transverse ray aberrations are derived and the QAC algorithm has been used to reduce their magnitudes for three test lenses, a 50mm f/4 triplet, a 90mm f/2 Petzval lens and a 75mm f/1 IR triplet. The results show that after DLS optimization of the higher order corrected designs using the default weighting factors, several different and useful solutions could be found. However, these alternative solutions were not necessarily superior designs and in some cases were significantly inferior. For example, in the higher order correction of a 50mm f/4 triplet, the starting design for which consisted of plane parallel plates separated by only 1mm air spaces, produced very steep curvatures to achieve the correction although the air spaces remained almost unchanged. This correction however was sufficient to guide the DLS algorithm towards a Gauss type objective with very steep curvatures about the stop and away from the superior triplet design.

The results presented in chapter 4 show that there does not appear to be any *consistent* advantage to be gained when higher order aberration correction is used to find starting designs for a standard gradient optimization algorithm. The use of this technique may only become apparent when further examples are tested.

In chapter 5, the adjustment of the final aberration balance based on maximizing the Strehl ratio of an optical system has been discussed and implemented. The method consists of evaluating the higher order aberration coefficients and calculating the third order aberrations which minimize the variance of the wavefront aberration. The third order aberrations can be adjusted using the QAC algorithm and provided the parameter changes are small, the higher order aberrations remain effectively static. This approach was tried on four test lenses and produced a significant increase in the spatial frequency response for two of the examples. One of the test lenses, a 90mm f/2 Petzval lens, was too far from the diffraction limit for there to be any significant improvement although the wavefront aberrations were noticeably different. The final test lens, a 25mm f/4 Erfle eyepiece, requested very large changes in the Seidel aberrations and so the underlying assumption broke down. This lens was close to the diffraction limit and it is still not clear why this situation arose.

The advantage of using the QAC algorithm in this context is that if one aberration is intractable then it does not prevent the other primary aberrations from reaching their target values. Since the total variance is simply the sum of the variances of the odd and even parts of the wavefront aberration then the minimum total variance is found. An intractable aberration at this stage of the design would not be uncommon since most aberrations would be somewhere close to their limiting values. If this intractable aberration remained in the change matrix then it would strongly influence the parameter changes and a worse solution is likely to result. A comparison with a gradient algorithm which uses the variance of the wavefront aberration as a merit function is necessary tc confirm this hypothesis.

There are many available methods for computing the wavefront aberration coeffi-

cients. The approach taken in chapter 5 is to form an exactly determined set of linear equations from the finite wavefront aberrations computed from a ray trace and the wavefront aberration polynomial coefficients expanded to the seventh order spherical aberration term. This requires nine rays to be traced and for only one off-axis field position. It would probably be more appropriate to use both off axis field positions and extend the number of terms in the approximating polynomial accordingly. It has recently been found that this is similar to the approach described by King [6.4], MacDonald [6.5] and other workers at Reading University although in this thesis the primary aberration coefficients are derived from the Seidel sums. Other methods include least squares fitting of the aberration polynomial to the wavefronts or the procedure described by Andersen [6.6]. The method described in this thesis is only appropriate for systems of moderate aperture and field specification and needs to be enhanced further if it is to be of more general applicability.

The major limitation with the Strehl ratio optimization is that the off axis field position for optimization must be specified by the user. A lens designer who only considers one off axis field position would be restricted to a very limited class of systems and so more sophisticated methods of balancing the aberrations with respect to field need to be investigated. At an elementary level, the performance could be averaged with respect to field but a more desirable option would probably be to use a weighted average, for example, it is unusual for a system to have as good a performance at the margin of the field as at the zone of the field. The balancing of aberrations with respect to field is a complex problem which has been outside the scope of this thesis and is probably worth further study in its own right. A more sophisticated approach to this problem may involve, for example, the use of through focus OTF data.

References

[6.1] D. Day, *Private communication*, 1989.

[6.2] C.G. Wynne, "Lens Designing by Electronic Digital Computer I," Proc. Phys. Soc., **73**, 777–787 (1959).

[6.3] E. Glatzel, "Ein neues Verfahren zur automatischen Korrektion optischer Systeme mit elektronischen Rechenmaschinen," Optik, **18**, 577–580 (1961).

[6.4] W.B. King, PhD Thesis, University of London, 1966.

[6.5] J. MacDonald, PhD Thesis, University of Reading, 1974.

[6.6] T.B. Andersen, "Automatic Computation of Optical Aberration Coefficients" Appl. Opt., **19**, 3800–3816 (1980).

Appendix A. Optimization Nomenclature and Definitions

For an objective function $f(\mathbf{x})$, the *gradient vector* $\nabla(f(\mathbf{x}))$ is the vector of first-order partial derivatives

$$\nabla(f(\mathbf{x})) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \tag{A.1}$$

and the $n \times n$ symmetric of second order derivatives is known as the *Hessian H*

$$H(f(\mathbf{x})) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}. \tag{A.2}$$

A function is *convex* if it is never underestimated by linear interpolation i.e. the chord joining the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lies on or above the graph of $f(x)$. A twice differentiable function is therefore only convex if

$$f''(x) \geq 0 \tag{A.3}$$

and hence a function of n variables is convex only if its matrix of second order partial differentials is positive semi-definite everywhere;

$$\mathbf{s}^{\mathrm{T}}\mathbf{A}\mathbf{s} \geq 0 \qquad \text{(for all vectors } \mathbf{s} \neq 0\text{)}. \tag{A.4}$$

The main importance of convexivity is that if $f(\mathbf{x})$ is a convex function in the convex region R then any local minimizer $\mathbf{x}^*$ of $f(\mathbf{x})$ in R is also a global minimizer.

References

[A.1] E.M.L. Beale, *Introduction to Optimization*, John Wiley & Sons, Chichester (1988).

[A.2] A.L. Peressini *et al.*, *The Mathematics of Non-linear Programming*, Springer-Verlag, New York, (1988).

[A.3] P.R. Adby and M.A.H. Dempster, *Introduction to Optimization Methods*, Chapman and Hall, London (1974).

Appendix B. Proof that $\mathbf{x} = \mathbf{A^T w}$.

This result follows from the following mathematical theorem for a minimum norm solution:

If A is an $m \times n$-matrix with linearly independent rows, $\mathbf{b} \in R^m$ and the linear system $A\mathbf{x} = \mathbf{b}$ is underdetermined, then a solution $\mathbf{x}^*$ is the solution that minimises the norm of the solution vector if and only is $\mathbf{x}^* \in \Re(A^T)$

From this since $\mathbf{x}^* \in \Re(A^T)$ (where $\Re$ is the range of the matrix $\mathbf{A^T}$ which is a subspace of R^m) and the m columns of A^T are simply the m linearly independent rows of A then $\mathbf{x}^* = w_1\mathbf{a_1} + w_2\mathbf{a_2} + \ldots + w_m\mathbf{a_m}$.

For further details of the mathematics of minimum norm solutions, see for example, [B.1].

References

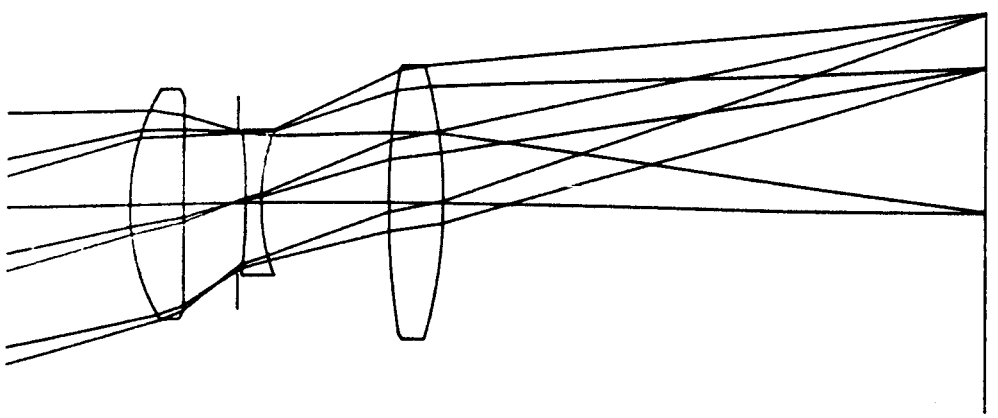
[B.1] A.L. Peressini *et al.*, *The Mathematics of Non-linear Programming*, Springer-Verlag, New York, (1988).

Appendix C – Test Lenses Used in Chapter 3.

TL3.1

EFL:- 48.237708 mm Mag:- 0.00 u'= -0.125
Lagrange Invariant = -1.5786 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

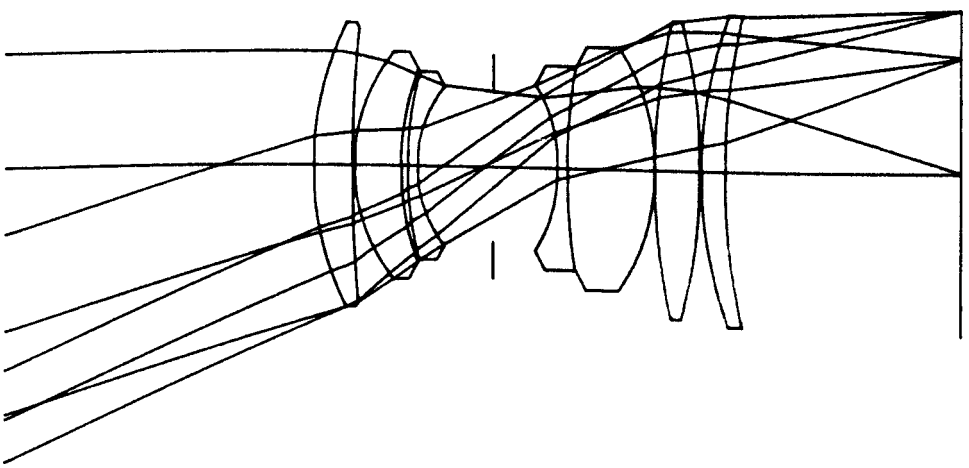
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.06819	14.66	0.000	1.00000	1.00000	1.00000	0.00000	9.18	Air
2	0.00141	707.71	3.500	1.69100	1.68716	1.69980	0.01263	9.03	LAK9
3	0.00000	plane	3.521	1.00000	1.00000	1.00000	0.00000	4.46	Air
4	-0.02932	-34.10	0.521	1.00000	1.00000	1.00000	0.00000	4.62	AIR
5	0.07609	13.14	1.000	1.68893	1.68250	1.70461	0.02211	4.42	SF8
6	0.01921	52.06	8.251	1.00000	1.00000	1.00000	0.00000	6.56	AIR
7	-0.03132	-31.92	3.500	1.69100	1.68716	1.69980	0.01263	7.07	LAK9
i	0.00000	plane	34.994	1.00000	1.00000	1.00000	0.00000	13.65	AIR



TL3.2

EFL:- 50.022650 mm Mag:- 0.00 u'= -0.294
Lagrange Invariant = -6.1603 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.02417	41.37	0.000	1.00000	1.00000	1.00000	0.00000	39.26	Air
2	0.00552	181.06	5.073	1.80801	1.80209	1.82200	0.01991	39.15	LASF3
3	0.04192	23.85	0.250	1.00000	1.00000	1.00000	0.00000	31.89	Air
4	0.02862	34.94	6.002	1.80801	1.80209	1.82200	0.01991	30.55	LASF3
5	0.02094	47.75	0.862	1.00000	1.00000	1.00000	0.00000	16373.96	Air
6	0.05907	16.93	1.250	1.78472	1.77598	1.80646	0.03048	24.62	SF11
7	0.00000	plane	9.759	1.00000	1.00000	1.00000	0.00000	9.64	Air
8	-0.04746	-21.07	8.259	1.00000	1.00000	1.00000	0.00000	23.79	Air
9	0.01261	79.28	1.250	1.78472	1.77598	1.80646	0.03048	27.88	SF11
10	-0.03487	-28.67	11.257	1.80801	1.80209	1.82200	0.01991	29.00	LASF3
11	0.01145	87.32	0.000	1.00000	1.00000	1.00000	0.00000	30.64	Air
12	-0.01278	-78.22	5.747	1.72000	1.71568	1.72997	0.01429	31.52	LAK10
13	0.01692	59.12	0.250	1.00000	1.00000	1.00000	0.00000	31.58	Air
14	0.01122	89.10	3.000	1.72000	1.71568	1.72997	0.01429	31.44	LAK10
i	0.00000	plane	30.341	1.00000	1.00000	1.00000	0.00000	53.95	Air

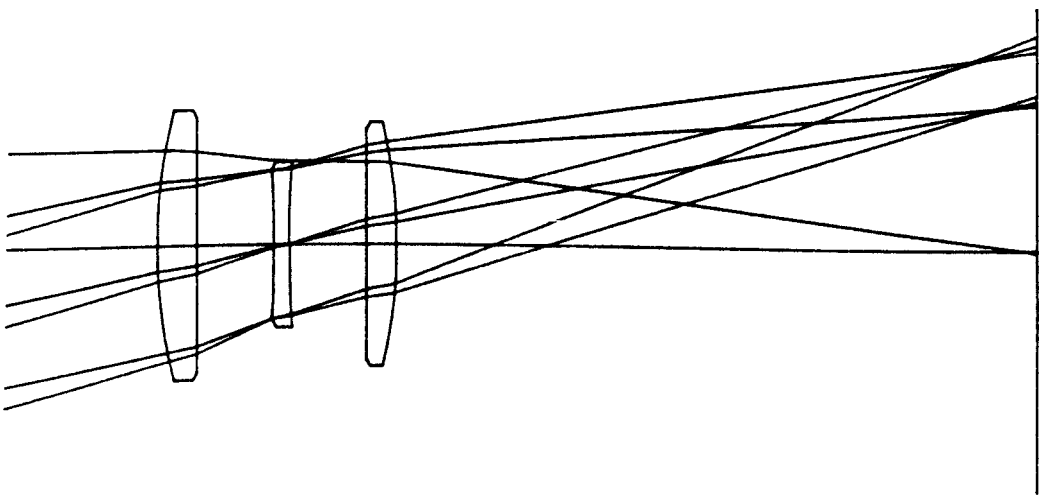


Appendix D – Test Lenses Used in Chapter 4.

TL4.1

EFL:- 49.165863 mm Mag:- 0.00 u'= -0.125
Lagrange Invariant = -1.6089 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

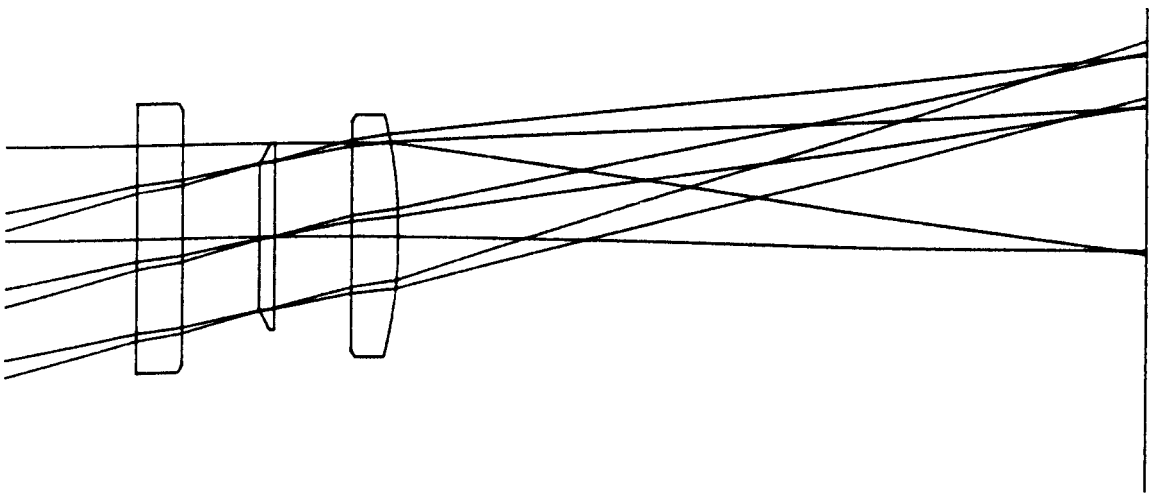
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.03000	33.33	0.000	1.00000	1.00000	1.00000	0.00000	9.04	Air
2	0.00000	plane	2.582	1.69100	1.68716	1.69980	0.01263	8.36	LAK9
3	-0.01500	-66.66	5.000	1.00000	1.00000	1.00000	0.00000	5.75	Air
4	0.01500	66.66	1.000	1.68893	1.68250	1.70461	0.02211	5.27	SF8
5	0.00000	plane	5.000	1.00000	1.00000	1.00000	0.00000	7.43	Air
6	-0.03000	-33.33	1.938	1.69100	1.68716	1.69980	0.01263	7.71	LAK9
i	0.00000	plane	41.284	1.00000	1.00000	1.00000	0.00000	14.00	Air



TL4.2

EFL:- 48.234234 mm Mag:- 0.00 u'= -0.125
Lagrange Invariant = -1.5785 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

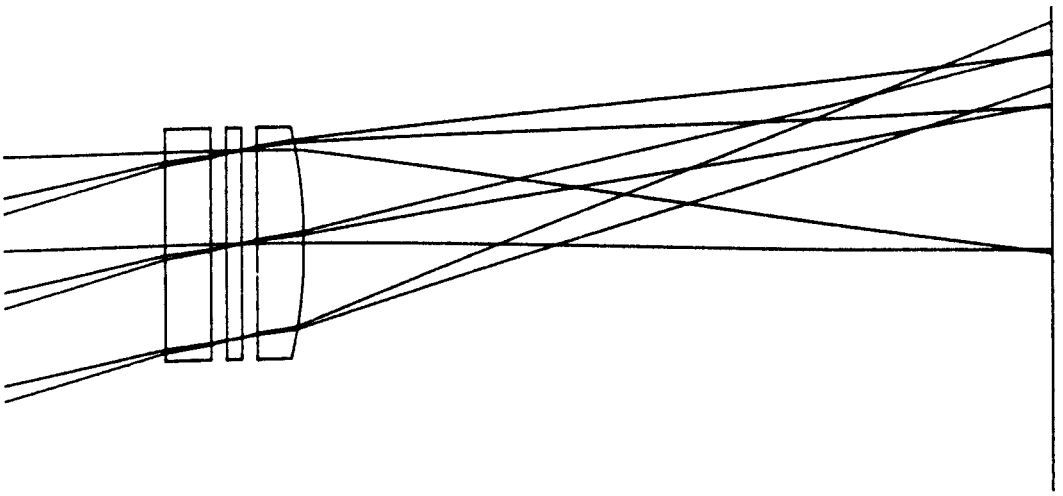
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.00000	plane	0.000	1.00000	1.00000	1.00000	0.00000	9.04	Air
2	0.00000	plane	3.000	1.69100	1.68716	1.69980	0.01263	8.36	LAK9
3	0.00000	plane	5.000	1.00000	1.00000	1.00000	0.00000	6.18	Air
4	0.00000	plane	1.000	1.68893	1.68250	1.70461	0.02211	6.03	SF8
5	0.00000	plane	5.000	1.00000	1.00000	1.00000	0.00000	7.43	Air
6	-0.03000	-33.33	3.000	1.69100	1.68716	1.69980	0.01263	7.71	LAK9
i	0.00000	plane	48.234	1.00000	1.00000	1.00000	0.00000	14.08	Air



TL4.3

EFL:- 48.234234 mm Mag:- 0.00 u'= -0.125
Lagrange Invariant = -1.5785 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

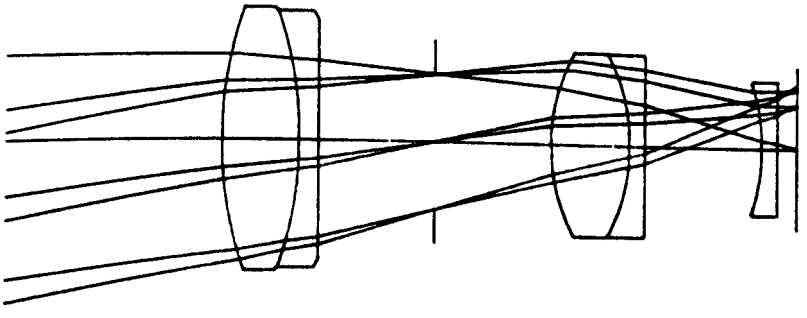
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.00000	plane	0.000	1.00000	1.00000	1.00000	0.00000	9.04	Air
2	0.00000	plane	3.000	1.69100	1.68716	1.69980	0.01263	8.36	LAK9
3	0.00000	plane	1.000	1.00000	1.00000	1.00000	0.00000	7.59	Air
4	0.00000	plane	1.000	1.68893	1.68250	1.70461	0.02211	6.03	SF8
5	0.00000	plane	1.000	1.00000	1.00000	1.00000	0.00000	7.43	Air
6	-0.03000	-33.33	3.000	1.69100	1.68716	1.69980	0.01263	7.71	LAK9
i	0.00000	plane	48.234	1.00000	1.00000	1.00000	0.00000	15.11	Air



TL4.4

EFL:- 87.469244 mm Mag:- 0.00 u'= -0.250
Lagrange Invariant = -3.8166 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

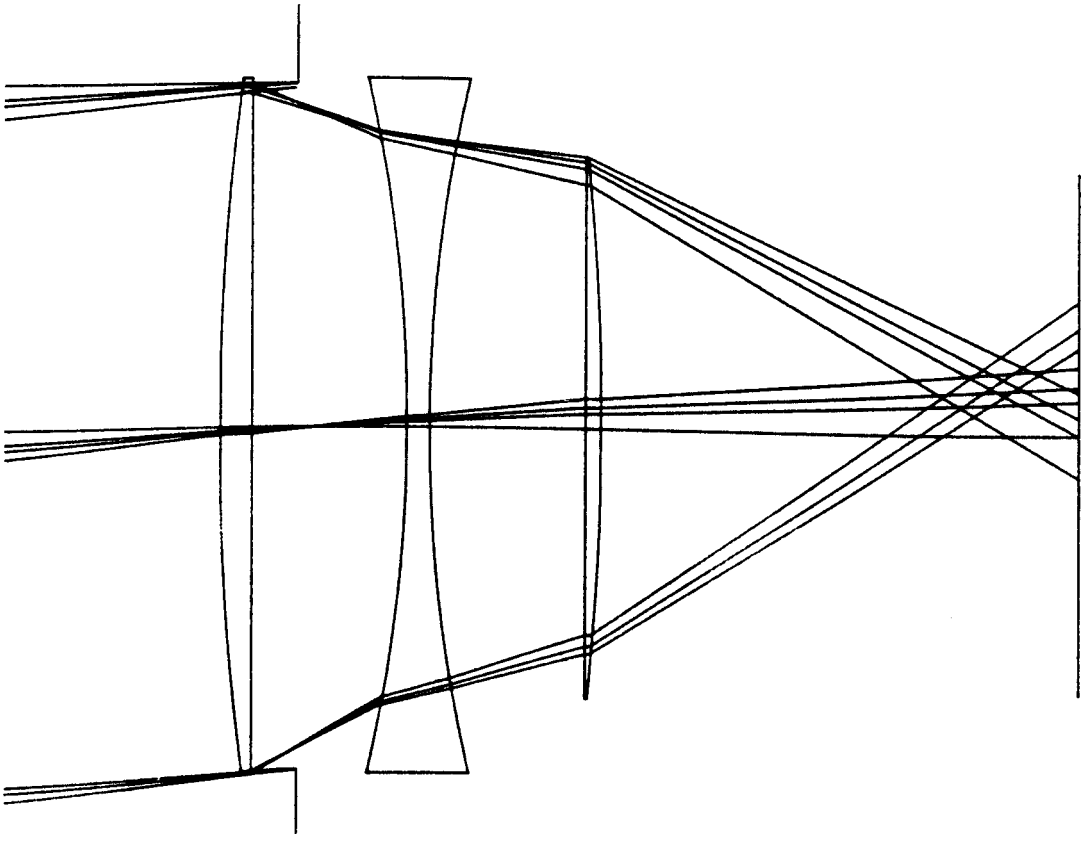
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.01000	100.00	0.000	1.00000	1.00000	1.00000	0.00000	33.42	Air
2	-0.01000	-100.00	20.000	1.58144	1.57723	1.59147	0.01424	32.44	LF5
3	0.00000	plane	5.000	1.74000	1.73241	1.75867	0.02625	30.24	SF3
4	0.00000	plane	30.000	1.00000	1.00000	1.00000	0.00000	17.14	Air
5	0.02000	50.00	30.000	1.00000	1.00000	1.00000	0.00000	23.57	Air
6	-0.02000	-50.00	20.000	1.68273	1.67816	1.69350	0.01535	23.38	BAF50
7	0.00000	plane	4.000	1.74000	1.73241	1.75867	0.02625	22.86	SF3
8	-0.02000	-50.00	30.000	1.00000	1.00000	1.00000	0.00000	16.58	Air
9	0.00000	plane	4.000	1.74000	1.73241	1.75867	0.02625	17.30	SF3
i	0.00000	plane	5.178	1.00000	1.00000	1.00000	0.00000	17.98	Air



TL4.5

EFL:- 88.651064 mm Mag:- 0.00 u'= -0.500
Lagrange Invariant = -4.4097 Defocus = -0.1871
Wavelength 1 = 10 microns Wavelength 2 = 12 microns Wavelength 3 = 8 microns

#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.00000	plane	0.000	1.00000	1.00000	1.00000	0.00000	44.33	Air
2	0.00300	333.33	-10.000	1.00000	1.00000	1.00000	0.00000	45.02	Air
3	0.00000	plane	4.200	4.00329	4.00202	4.00551	0.00349	44.87	Ge
4	-0.00500	-200.00	20.000	1.00000	1.00000	1.00000	0.00000	38.47	Air
5	0.00500	200.00	3.000	1.49470	1.48000	1.50640	0.02640	37.14	NaCl
6	0.00000	plane	20.000	1.00000	1.00000	1.00000	0.00000	35.47	Air
7	-0.00300	-333.33	2.000	4.00329	4.00202	4.00551	0.00349	35.46	Ge
i	0.00000	plane	61.253	1.00000	1.00000	1.00000	0.00000	17.24	Air



Appendix E. Proof that $Var(W) = Var(W_{even}) + Var(W_{odd})$.

Central to the work presented in chapter 5 is the assumption that the variance of the total wavefront aberration, $Var(W)$, is equal to the sum of the variances of several even and odd contributions to the wavefront aberration. In this appendix, a proof of this identity is provided.

Define a set of even functions $f_1(x), f_2(x), \ldots, f_m(x)$ on the interval $-1 \leq x \leq 1$ and a set of odd functions $g_1(x), g_2(x), \ldots, g_n(x)$ also on the interval $-1 \leq x \leq 1$. The variance of the even contributions to the total variance, $Var(f)$, is given by

$$\begin{aligned} Var(f) &= \int_{-1}^{1} (f_1 + f_2 + \ldots)^2 dx - \left\{ \int_{-1}^{1} (f_1 + f_2 + \ldots) dx \right\}^2 \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\int_{-1}^{1} f_i f_j dx \right) - \left\{ \sum_{i=1}^{m} \int_{-1}^{1} f_i dx \right\}^2 \end{aligned} \tag{E.1}$$

and the variance of the odd contributions to the total variance by

$$\begin{aligned} Var(g) &= \int_{-1}^{1} (g_1 + g_2 + \ldots)^2 dx - \left\{ \int_{-1}^{1} (g_1 + g_2 + \ldots) dx \right\}^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \int_{-1}^{1} g_i dx \end{aligned} \tag{E.2}$$

A value for the total variance, $Var'(W)$, can be found by summing equations (E.1) and (E.2) giving that

$$Var'(W) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\int_{-1}^{1} f_i f_j dx \right) + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\int_{-1}^{1} g_i g_j dx \right) - \left\{ \sum_{i=1}^{m} \int_{-1}^{1} f_i dx \right\}^2 \tag{E.3}$$

Finally define the function W to be the sum of its even and odd contributions,

$$W = f_1 + f_2 + \ldots + g_1 + g_2 + \ldots$$

then the total variance of W is given by

$$Var(W) = \int_{-1}^{1} (f_1 + f_2 + g_1 + g_2 + \ldots)^2 dx - \left\{ \int_{-1}^{1} (f_1 + f_2 + g_1 + g_2 + \ldots) dx \right\}^2$$

If the squared bracket which forms the integrand of the first term is multiplied out before performing the integration all the cross terms of odd and even functions are zero. Similarly, the integrals of the odd contributions in the second term are also zero leaving

$$Var(W) = \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\int_{-1}^{1} f_i f_j dx\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\int_{-1}^{1} g_i g_j dx\right) - \left\{\sum_{i=1}^{m}\int_{-1}^{1} f_i dx\right\}^2 \quad (E.4)$$

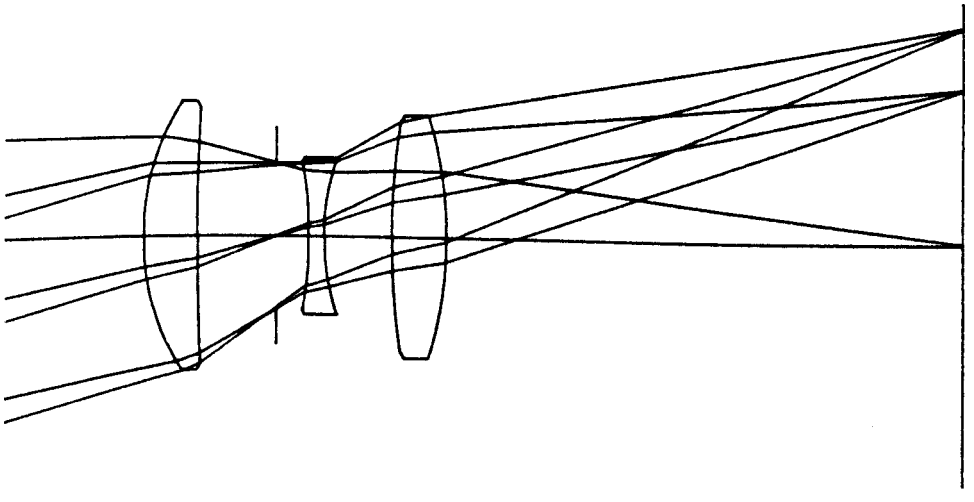
which is identical to equation (E.3).

Appendix F – Test Lenses Used in Chapter 5.

TL5.1

EFL:- 51.040430 mm Mag:- 0.00 u'= -0.125
Lagrange Invariant = -1.6703 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

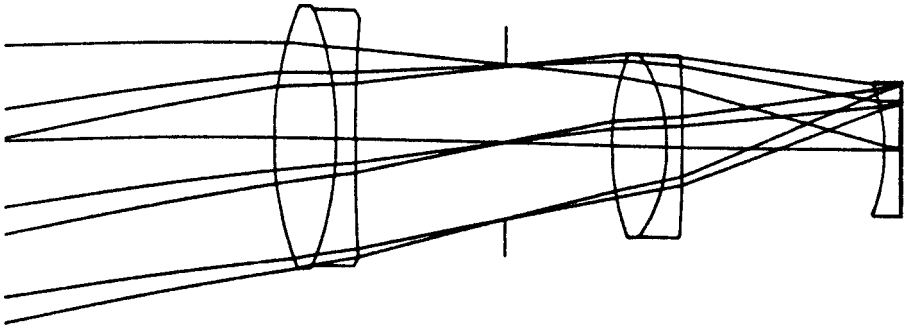
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.06116	16.35	0.000	1.00000	1.00000	1.00000	0.00000	8.70	AIR
2	0.00641	156.03	3.500	1.69100	1.68716	1.69980	0.01263	8.20	LAK9
3	0.00000	plane	5.101	1.00000	1.00000	1.00000	0.00000	4.58	AIR
4	-0.04008	-24.95	2.101	1.00000	1.00000	1.00000	0.00000	4.74	AIR
5	0.06386	15.66	1.000	1.68893	1.68250	1.70461	0.02211	5.05	SF8
6	0.01978	50.56	4.367	1.00000	1.00000	1.00000	0.00000	7.44	AIR
7	-0.03896	-25.67	3.500	1.69100	1.68716	1.69980	0.01263	7.85	LAK9
i	0.00000	plane	33.221	1.00000	1.00000	1.00000	0.00000	14.01	AIR



TL5.2

EFL:- 97.656003 mm Mag:- 0.00 u'= -0.250
Lagrange Invariant = -4.2610 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

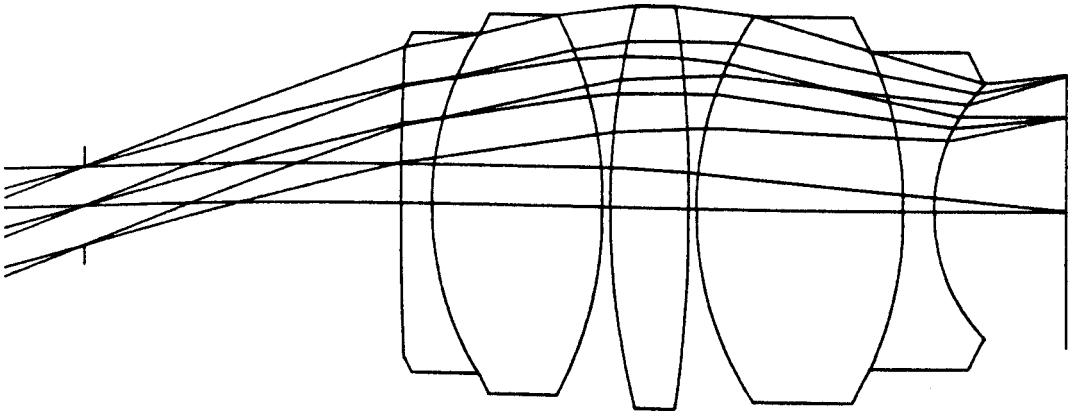
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.01092	91.58	0.000	1.00000	1.00000	1.00000	0.00000	33.42	Air
2	-0.01067	-93.76	16.000	1.58144	1.57723	1.59147	0.01424	32.44	LF5
3	0.00203	491.88	5.000	1.74000	1.73241	1.75867	0.02625	30.24	SF3
4	0.00000	plane	38.707	1.00000	1.00000	1.00000	0.00000	19.32	Air
5	0.01503	66.56	27.207	1.00000	1.00000	1.00000	0.00000	23.57	Air
6	-0.02370	-42.19	14.000	1.68273	1.67816	1.69350	0.01535	23.38	BAF50
7	-0.00234	-426.80	4.000	1.74000	1.73241	1.75867	0.02625	22.86	SF3
8	-0.02192	-45.63	52.000	1.00000	1.00000	1.00000	0.00000	16.58	Air
9	0.00191	524.11	4.000	1.74000	1.73241	1.75867	0.02625	17.30	SF3
i	0.00000	plane	0.677	1.00000	1.00000	1.00000	0.00000	17.38	Air



TL5.3

EFL:- 25.061567 mm Mag:- 0.00 u'= -0.100
Lagrange Invariant = -0.8748 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

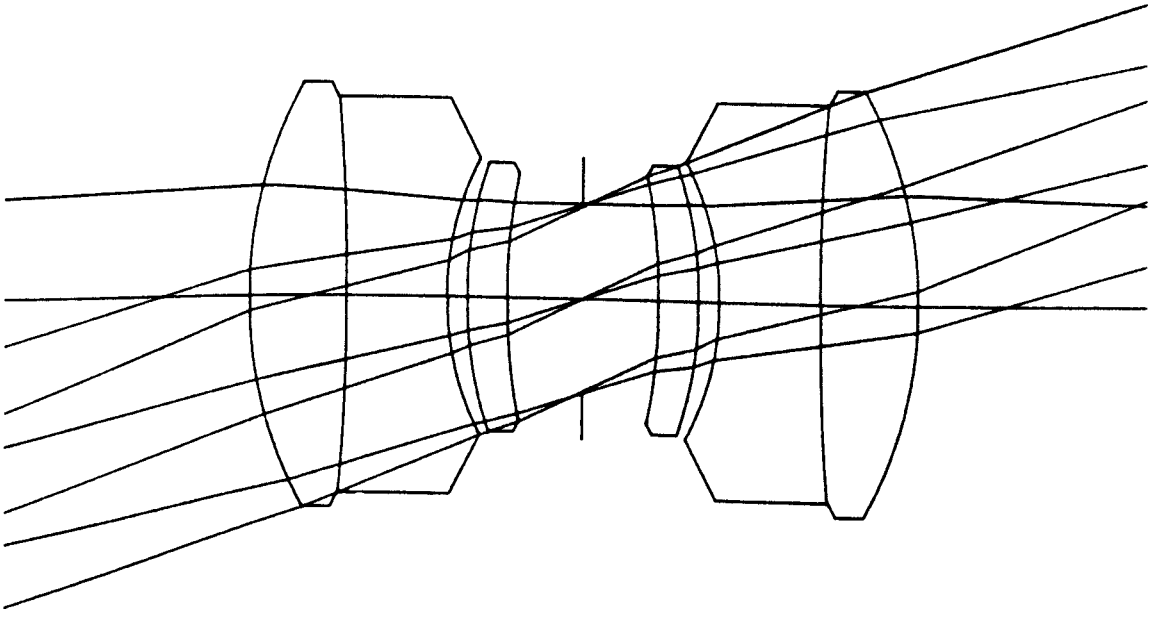
#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.00000	plane	0.000	1.00000	1.00000	1.00000	0.00000	2.51	Air
2	0.00432	231.54	20.500	1.00000	1.00000	1.00000	0.00000	10.05	Air
3	0.04641	21.55	2.000	1.61659	1.61164	1.62848	0.01684	11.00	F4
4	-0.03700	-27.02	11.000	1.51680	1.51432	1.52238	0.00806	12.31	BK7
5	0.01886	53.02	0.500	1.00000	1.00000	1.00000	0.00000	13.04	Air
6	-0.01066	-93.83	5.000	1.61765	1.61427	1.62547	0.01121	13.01	SSK4
7	0.04356	22.96	0.500	1.00000	1.00000	1.00000	0.00000	12.46	Air
8	-0.03932	-25.43	13.300	1.60881	1.60567	1.61600	0.01034	10.25	SK3
9	0.08322	12.02	2.000	1.64831	1.64271	1.66187	0.01916	8.20	SF12
i	0.00000	plane	8.478	1.00000	1.00000	1.00000	0.00000	8.84	Air



TL5.4

EFL:- 99.766307 mm Mag:- -1.00 u'= 0.000
Lagrange Invariant = -2.8796 Defocus = 0.0000
Wavelength 1 = 587.6nm Wavelength 2 = 656.3nm Wavelength 3 = 486.1nm

#	curve	radius	sep.	n1	n2	n3	dsp	clrad	glass
1	0.03440	29.07	177.537	1.00000	1.00000	1.00000	0.00000	13.94	Air
2	-0.00635	-157.53	6.300	1.62041	1.61727	1.62756	0.01029	13.05	SK16
3	0.05091	19.64	6.500	1.54814	1.54456	1.55655	0.01199	9.06	LLF1
4	0.03674	27.22	1.300	1.00000	1.00000	1.00000	0.00000	8.84	Air
5	0.02392	41.80	2.600	1.69100	1.68716	1.69980	0.01263	8.14	LAK9
6	0.00000	plane	4.872	1.00000	1.00000	1.00000	0.00000	6.11	Air
7	-0.02392	-41.80	4.872	1.00000	1.00000	1.00000	0.00000	8.14	Air
8	-0.03674	-27.22	2.600	1.69100	1.68716	1.69980	0.01263	8.85	LAK9
9	-0.05091	-19.64	1.300	1.00000	1.00000	1.00000	0.00000	9.06	Air
10	0.00635	157.53	6.500	1.54814	1.54456	1.55655	0.01199	13.06	LLF1
11	-0.03440	-29.07	6.300	1.62041	1.61727	1.62756	0.01029	13.94	SK16
i	0.00000	plane	177.506	1.00000	1.00000	1.00000	0.00000	73.70	Air



REFERENCES

P.R. Adby and M.A.H. Dempster, *Introduction to Optimization Methods*, Chapman and Hall, London (1974), p 103. [2.21], [A.3]

T.B. Andersen, "Automatic Computation of Optical Aberration Coefficients" Appl. Opt., **19**, 3800–3816 (1980). [6.6]

K.E. Atkinson, *Introduction to Numerical Analysis*, John Wiley, New York (1978) 450–454. [3.11]

J.G. Baker, Technical Reports 1–12 and a Supplement prepared by Perkin-Elmer Corp., under Air Force Contract AF 33(038)–10836. Declassified 1959. [1.1]

R. Barakat and A. Houston, "Transfer Function of an Optical System in the Presence of Off-Axis Aberrations," J.O.S.A., **55**, 1142–1148 (1965). [5.20]

E.M.L. Beale, *Introduction to Optimization*, John Wiley & Sons, Chichester (1988). [A.1]

W. Bessematter, "A Sophisticated Optical Design Program for Microcomputers," in Proc. SPIE Vol. **399**, *Optical System Design, Analysis and Production*, 66–73 (1983) [2.49]

G. Black, "On the Automatic Design of Optical Systems," Proc. Phys. Soc, B, **68**, 729–736 (1955). [2.3], [1.2]

I. Bohachevsky, V.K. Viswanathan, and G. Woodfin, "An 'Intelligent' Optical Design Program," in Proc. SPIE Vol. **485**, *Applications of Artificial Intelligence*, 104–112 (1984) [1.23], [2.50]

M. Born, and E. Wolf, *Principles of Optics*, Pergamon Press, 1966, p 461. [5.18]

B. Brixner, "Lens Design Merit Functions: rms image spot size and rms optical path difference," Appl. Opt., **17**, 715–716 (1978). [5.9]

D.R. Buchele, "Damping Factor for the Least-squares Method of Optical Design," Appl. Opt., **7**, 2433–2435 (1968). [1.15], [2.24]

A. Cauchy, Comptes Rendus, **25**, 536–538 (1847). [2.5]

A. Choudry, and R.B. Johnson, "Knowledge Based Environment for Optical System Design," in Proc. SPIE Vol. 805, *Optical Components and Systems,* (1987). [1.25], [2.64]

A.E. Conrady, *Applied Optics and Optical Design*, Dover, New York (1960), p585. [5.1]

A. Cox, *A System of Optical Design*, Focal Press, London (1964). [1.28], [4.2]

L.W. Cornwell *et al.*, "Grey's Method for Non-linear Optimization," J.O.S.A., **63,** 576–581 (1973). [2.39]

F.D. Cruickshank, "A System of Transfer Coefficients for Use in the Design of Lens Systems," Proc. Phys. Soc., **57**, 350–361 (1945). [3.5]

D. Day, *Private communication*, 1989. [6.1]

D.C. Dilworth, "Pseudo-second Derivative Matrix and its Application to Lens Design," Appl. Opt., **17,** 3372–3375 (1978). [1.17], [2.30]

D.C. Dilworth, "Automatic Lens Optimization: recent improvements," in Proc. SPIE Vol. **554,** *International Lens Design Conference,* 191–196 (1985). [2.31]

D.C. Dilworth, "Applications of Artificial Intelligence to Computer-Aided Lens Design," in Proc. SPIE Vol. **766,** *Recent Trends in Optical System Design: Computer Lens Design Workshop,* 91–99 (1987). [1.24], [2.61]

J. Dunning-Davies, *Mathematical Methods for Mathematicians, Physical Scientists and Engineers,* Ellis-Horwood (1982), 186–187. [2.38]

A. Faggiano, "Automatic Lens Design with Pseudo-second-derivative Matrix. A Contribution." Appl. Opt., **19,** 4226–4229 (1980). [2.32]

D.P. Feder, "Automatic Lens Design Methods," JOSA, **47**, 902–912 (1957). [2.6],[3.14]

D.P. Feder, "Automatic Optical Design," Appl. Opt., **2**, 1209–1226 (1963). [2.10]

D.P. Feder, "Automatic Lens Design with a High-speed Computer," JOSA, **52**, 172–176 (1962). [1.7], [2.13]

D.P. Feder, "Automatic Lens Design Methods," J.O.S.A., **47,** 902–912 (1957). [1.5], [2.6]

R. Fletcher, "A Modified Marquardt Subroutine for Non-linear Least-squares," A.E.R.E. (UK) Harwell Technical Report No. R6799 (1971). [2.20]

G.W. Forbes and A.E.W. Jones, "Towards Global Optimization with Adaptive Simulated Annealing", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, 144–153 (1990). [1.20], [2.56]

G.W. Forbes and A.E.W. Jones, "Global Optimization in Lens Design", Optics and Photonics News, 23–29, March 1992. [1.21]

E. Glatzel, "Ein neues Verfahren zur automatischen Korrektion optischer Systeme mit elektronischen Rechenmaschinen," Optik, **18**, 577–580 (1961). [1.11], [2.40], [6.3]

E. Glatzel, "A New Method for the Automatic Correction of Optical Systems Using Electronic Computers", in Proc. Conf. *Lens Design with Large Computers,* Rochester, NY, USA, 5–8 July 1966, 23-41—23-47. [1.12]

E. Glatzel and R. Wilson, "Adaptive Automatic Correction in Optical Design," Appl. Opt., **7**, 265–276 (1968). [1.26], [2.44], [3.2]

R.W. Gostick, "OTF-based Optimization Criteria for Automatic Optical Design", Opt. & Quant. Elec., **8**, 31–37 (1976). [5.12]

Gradsteyn and Ryzhik, *Tables of Integrals, Series and Products*, 0.319, p12, Academic Press (1986). [3.3]

D.S. Grey, "Recent Developments in Orthonormalization of Parameter Space," in Proc. Conf. *Lens Design with Large Computers,* Rochester, NY, USA, 5–8 July 1966, 27-1—27-22. [2.2]

D.S. Grey, "Aberration Theories for Semi-automatic Lens Design by Electronic Computers I. Preliminary Remarks," JOSA, **53**, 672–676 (1963). [1.14], [2.28]

D.S. Grey, "Aberration Theories for Semi-automatic Lens Design by Electronic Computers II. A Specific Program," JOSA, **53**, 677–680 (1963). [2.37]

D.S. Grey, "Orthogonal Polynomials as Lens Aberration Coefficients," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, 85–90 (1980). [5.3]

T.I. Harris and M.J. Hayford, "New Developments in Code V Optimization," in Proc. SPIE Vol. **237**, *International Lens Design Conference,* 11–17 (1980). [2.17]

G.K. Hearn, " Generalised Simulated Annealing Optimization Used in Conjunction with Damped Least-squares Techniques" in Proc. SPIE Vol. **766**, *Recent Trends in Optical System Design*, 283–284 (1987). [1.18], [2.52]

G.K. Hearn, "Design Optimization Using Generalised Simulated Annealing," in Proc. SPIE Vol. **818**, *Current Developments in Optical Engineering II*, 258–264 (1987). [2.53]

G.K. Hearn, "Automation of Cost Control During Optimization," in Proc. SPIE Vol. **1049** (1989). [2.55]

M.R. Hestenes and E. Stiefel, J. Res. Nat. Bur. Std., **49**, 409–436 (1952). [2.12]

H.H. Hopkins and A. Kadkly, "Control of Damping Factors and Boundary Conditions in Optimization Programs," J. Mod. Opt., **35**, 49–74 (1988). [2.27]

H.H. Hopkins, "The use of diffraction-based criteria of image quality in automatic optical design," Opt. Act., **13**, 343–369 (1966). [5.6]

R.E. Hopkins *et al.*, "Automatic Correction of Third-Order Aberrations ", J.O.S.A., **45**, 363–365 (1955). [2.42], [1.3]

R.E. Hopkins and G.H. Spencer, "Creative Thinking and Computing Machines in Optical Design," J.O.S.A., **52**, 172–176 (1962). [1.13], [2.45]

E.D. Huber, "Extrapolated Least-squares Optimization in Optical Design," J.O.S.A. A, **2**, 544-554 (1985). [2.25],[3.7]

E.D. Huber, "Extrapolated least-squares optimization: a new approach to least-squares optimization in optical design," App. Op. **21**, 1705–1707 (1982). [2.34]

T.H. Jamieson, "Optimization Techniques in Lens Design," Adam Hilger Ltd., London (1971), p 19. [2.23]

M.J. Kidger and C.G. Wynne, "Experiments with Lens Optimization Procedures," Opt. Acta, **14**, 279–288 (1967). [2.19]

M.J. Kidger, PhD Thesis, University of London (1971). [2.26]

M.J. Kidger and P. Benham, "Optimization of the Optical Transfer Function," in Proc. SPIE Vol. **1354**, *International Lens Design Conference* , 92–96 (1990). [5.13]

W.B. King, PhD Thesis, University of London, (1966). [1.29], [6.4]

W.B. King, "Use of the Modulation-Transfer Function (MTF) as an Aberration-Balancing Merit Function in Automatic Lens Design," J.O.S.A., **59**, 1155–1158 (1969). [1.16], [5.5]

W.B. King, "Dependence of the Strehl Ratio on the Magnitude of the Variance of the Wave Aberration," J.O.S.A., **58**, 655–661 (1968). [5.10]

W.B. King, "Correlation between the Relative Modulation Function and the Magnitude of the Variance of the Wave-Aberration Difference Function,"J.O.S.A., **59,** 692–697 (1969). [5.11]

M. Krautter, "Zusammenhänge der in Programmen zur automatischen Korrektion verwendeten Lösungmatrizen," Optik, **30**, 332–337 (1970). [2.47]

T.G. Kuper and T.I. Harris, "A New Look at Global Optimization for Optical Design," Photonics Spectra, January 1992. [2.58]

K. Levenberg, "A Method for the Solution of Certain Non-linear Problems in Least-squares," Quart. J. Appl. Math., **2**, 164 (1944). [1.10], [2.16]

J. MacDonald, PhD Thesis, University of Reading, (1974). [6.5]

J. MacDonald and Y. Wang, "Analytical Derivatives in Damped Least-squares Optimization," in Proc. SPIE Vol. **655**, *Optical System Design, Analysis and Production for Advanced Technology Systems,* 49–53 (1986). [2.36],[3.6]

A. Maréchal, "Études des effets combines de la diffraction et des aberrations géométriques sur l'image d'un point lumineux," Rev. d'Opt., **27**, 257–277 (1947). [5.19]

D.W. Marquardt, "An Algorithm for Least-squares Estimation of Nonlinear Parameters," J. SIAM, **11**, 431–441 (1963). [2.18]

A.L. M'Aulay and F.D. Cruickshank, "A Differential Method of Adjusting the Aberrations of a Lens System," Proc. Phys. Soc., **57**, 302–310 (1945). [3.4]

J. Maxwell and C.C. Hull, "A Simple Multidimensional Quadratic Extrapolation Method for the Specific Correction of Aberrations in Optical Systems," in Proc. SPIE Vol. **1354**, *International Lens Design Conference* , 277–285 (1990). [3.1]

J. Maxwell and C.C. Hull, "A Multidimensional Quadratic Extrapolation Method for the Specific Correction of Aberrations in Optical Systems," Appl. Opt., **31**, 2234–2240 (1992). [1.27]

C. McCarthy, "A Note on the Correction of Third-order Aberrations", J.O.S.A., **45**, 1087–1088 (1955). [2.43]

J. Meiron and H.M. Loebenstein, "Automatic Correction of Residual Aberrations," JOSA, **47**, 1104–1109 (1957). [1.6], [2.8]

J. Meiron and G. Volinez, "Parabolic Approximation for Automatic Lens Design," JOSA, **50**, 207–211 (1960). [2.11], [1.4]

J. Meiron, "Automatic Lens Design by the Least-squares Method," J.O.S.A., **49**, 293–298 (1959). [2.15]

J. Meiron, "Damped Least-squares Method for Automatic Lens Design," J.O.S.A., **55**, 1105–1109 (1965). [2.22]

J. Meiron, "The use of Merit Functions based on Wavefront Aberrations in Automatic Lens Design," Appl. Opt., **7**, 667–672 (1968). [5.7]

G.H. Moss, "An Expert System for Preliminary Optical Design," in Proc. SPIE Vol. 1013, *Optical Design Methods, Applications and Large Optics,* (1988). [2.62]

D. Nicholls, *Private Communication*, 1989. [2.4]

A.L. Peressini *et al.*, *The Mathematics of Non-Linear Programming*, Springer-Verlag, New York (1980) 145–148. [2.41], [3.9], [A.2], [B.1]

W.H. Press *et al.*, *Numerical Recipes*, Cambridge University Press, Cambridge (1986), Chap. 2. [3.10], [3.12]

J.L. Rayces, "Ten Years of Design with Glatzel's Adaptive Method," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, 75–84 (1980). [2.48], [3.16], [5.2]

P.N. Robb, "Accelerating Convergence in Automatic Lens Design," Appl. Opt., **18**, 4191–4194 (1979). [2.33]

P.N. Robb, "Lens Design using Optical Aberration Coefficients," in Proc. SPIE Vol. **237**, *International Lens Design Conference*, 109–118 (1980). [5.8]

P.J. Rogers, *Private Communication*, (1991). [4.1]

Rong-Seng Chang, "Analytical Lens Design by Microcomputer with Artificial Intelligence," Proc. SPIE Vol. **554**, *International Lens Design Conference,* (1985). [2.59]

S. Rosen and C. Eldert, "Least-squares Method for Optical Correction," J.O.S.A., **44**, 250–252 (1954). [1.8], [2.14]

G.H. Spencer, "A Flexible Automatic Lens Correction Procedure," Appl. Opt., **2**, 1257–1264 (1963). [2.46]

D. Sturlessi, and D.C. O'Shea, "Future of Global Optimization in Optical Design", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, 54–68 (1990). [1.22], [2.57]

T. Takahashi, "An Experimental Analysis of Optimization Algorithms Using a Model Function," Optik, **35,** 101–115 (1972). [2.1]

B. Tatian, "Nonlinearity and Lens Design", in Proc. SPIE Vol. **1354**, *International Lens Design Conference*, 154–164 (1990). [2.35], [3.15]

B. Tatian, "Aberration Balancing Criterion for non diffraction-limited lenses," in Proc. SPIE Vol. **237**, *International Lens Design Conference* , 92–98 (1980). [5.14]

D. Veerabhadra Rao, and V. Venkateswara Rao, "Optimization Techniques in Lens Design," J. Sc. and Ind. Res. (India), **40**, 5, 297–312 (1981). [2.9]

V.K. Viswanathan *et al.*, "An Attempt to Develop an 'Intelligent' Lens Design Program," Proc. SPIE Vol. 554, *International Lens Design Conference,* 10–17 (1985). [2.51]

W.T. Welford, *Aberrations of Optical Systems,* Adam Hilger Ltd. (1986), p265. [4.3],[5.16]

S.W. Weller, "Simulated Annealing: What good is it?" in Proc. SPIE Vol. **818**, *Current Developments in Optical Engineering II*, 265–274 (1987). [1.19], [2.54]

S.W. Weller, "Optics Toolbox: an intelligent relational database system for optical designers," in Proc. SPIE Vol. **679**, *Current Developments in Optical Engineering and Diffraction Phenomena,* 7–11 (1986). [2.60]

C.S. Williams and O.A. Becklund, *Introduction to the Optical Transfer Function*, John Wiley & Sons, New York (1989), p131. [5.4]

C.J. Woodruff, "A Comparison using Orthogonal Coefficients, of two forms of Aberration Balancing," Opt. Act., **22,** 933–941 (1975). [5.15]

C.G. Wynne, "Lens Designing by Electronic Digital Computer I," Proc. Phys. Soc., **73**, 777–787 (1959). [1.9], [2.7], [6.2]

C.G. Wynne and P.M.J.H. Wormell, "Lens Design by Computer," Appl. Opt., **2,** 1233–1238 (1963). [2.29]