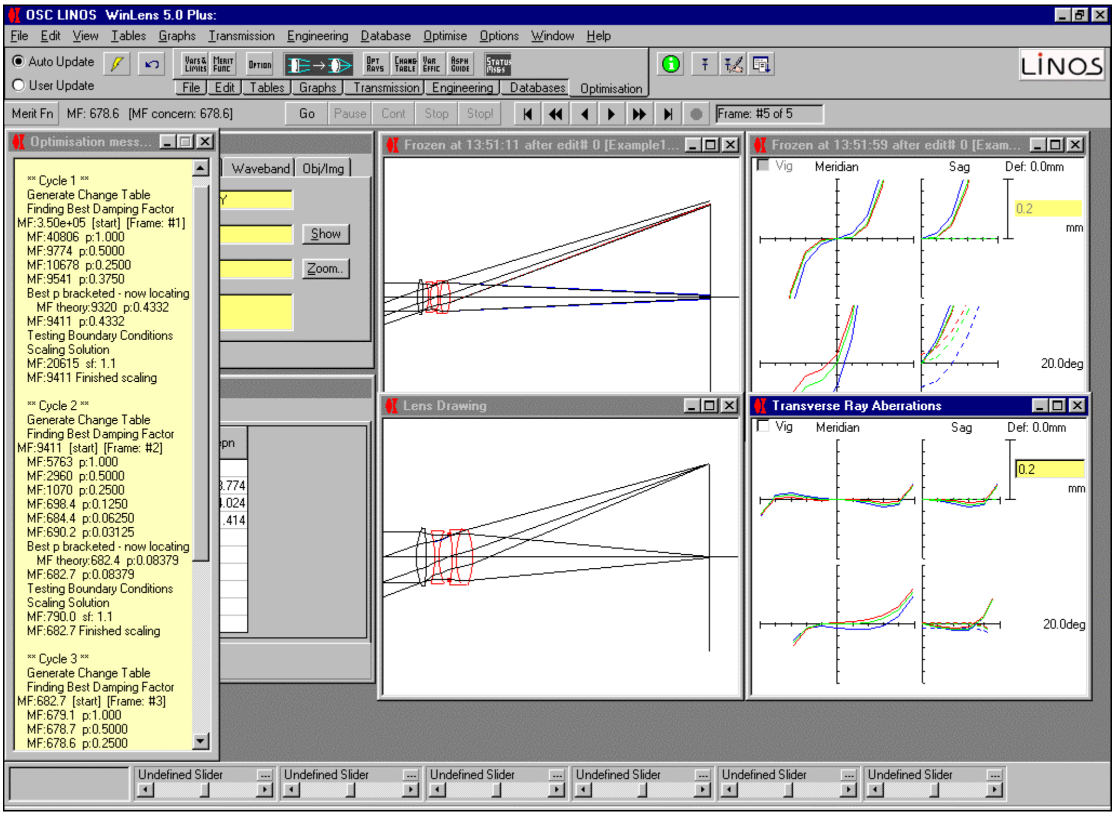
WinLens

Optical Design Software

users guide - optimisation



A Product of

LINOS Photonics GmbH
Königsallee 23
D - 37081 Göttingen
Germany

www.linos.de / www.winlens.de

Summary

This manual provides a comprehensive introduction to optimisation using WinLens. It contains:

- a simple overview of the manual
- a simple overview of optimisation features in Winlens
- detailed notes on all relevant tools
- a simple worked example
- notes on more complex issues
- an introduction to optimisation theory

Cross references are shown in ***Bold-Italics***. If you are reading this document on a PC, such cross references are hyperlinked to the relevant chapter or section.

This manual assumes that the user:

- has a modest knowledge of optics [Note: the WinLens help system contains a lot of background information]
- is reasonably familiar with Windows
- has had some experience with WinLens in the past – in particular creating lens files and saving/reading them from/to disk

Contents

Revision 1.1.8 *17th April 2002*

Author: Dr Geoff Adams

1 Introduction

- *What is optimisation*
- *Preparations before optimisation*
- *Quick guide to the manual*

"Science is a wonderful thing if one does not have to earn one's living from it."

Albert Einstein

1.1 What is optimisation

An optical system is defined by a set of parameters, such as radius, separation, glass etc. During optimisation the software alters some of these parameters, in order to improve the performance of the system. The performance is measured by a number of different aberrations/constraints or defects. These defects are combined into a merit function, which yields a single number. The smaller that number is, the better the system.

1.2 Preparations before optimisation

In order to optimise a system you therefore need the following:

- starting system
- variables
- merit function

In this manual we assume that you are used to creating or importing optical systems in WinLens.

1.3 Quick guide to the manual

The manual is organised as follows:

Chapter	Purpose
2: Quick guide	Shows the location of the various tools along with brief notes on their functions
3: Variables	How to use the variable editor to select variables and [optionally] set upper & lower limits on those variables. These hard limits cannot be violated in optimisation.
4: Merit function	Defines a merit function and discusses the WinLens smart merit function. • Quick guide to the merit function editor • The setup wizards & creating a merit function • Adding extra terms to the merit function • Editing the merit function
5: Optimisation options	Covers items such as ending optimisation, optimisation strategy, damping factor etc
6: Control during optimisation	How to pause/continue/abort an optimisation
7: Playback after optimisation	Lens files are saved at key stages during optimisation. This chapter shows how to replay the optimisation with the 'video' controls, for review and restart purposes.
8: How to	Discussion of some important or advanced topics: • Glass optimisation • Edge thickness control • Choose of variables • Choice of surfaces for asphterisation • Specific systems [zoom, aspherics etc]
9: Example	Worked examples with many illustrations
10: Glossary	Definitions of useful terms
11: Defects	Definitions of all defects available for use in the merit function
12: Appendix	Basics of optimisation theory
13: Index	

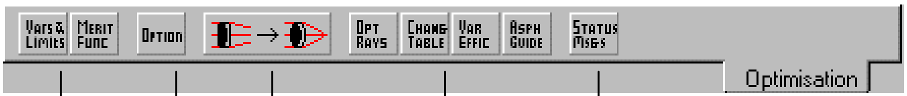
2 Quick guide to optimisation in WinLens

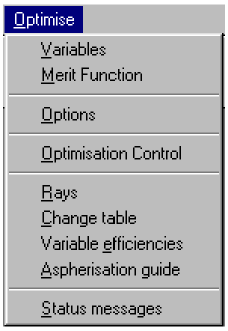
- *Location of the optimisation tools*
- *Brief description of their functions*
- *Cross references to full discussions*

"The perfect computer has been developed. You just feed in your problems and they never come out again."

Al Goodman

All optimisation tools can be accessed through the optimisation menu or through the optimisation tab:





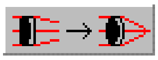
Editors for:
- variables ***Ch3***
- merit function ***Ch4***

Show/hide optimisation control toolbar

Status messages generated during optimisation

Optimisation options

Various tools to help in make choices for optimisation



To actually run an optimisation you need to display the optimisation toolbar. Select the 'Optimisation control' option from the menu or click on the large icon shown left.

The optimisation control toolbar [shown below] appears under the main toolbar.

Update merit function

Start/stop/pause & continue optimisation [***Ch6***]

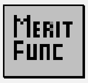
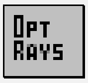
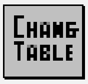
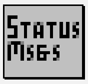
'Video' controls to replay optimisation (frames are # of systems saved during optimisation and ready for replay) [***Ch7***]

Once you have:
- selected variables [***Ch3***]
- defined a merit function [***Ch4***]

then press '**Go**' to start the optimisation. [simple examples are given in ***Ch 9***]

The optimisation will run until one of the following occur:
- All defects are within tolerance
- A minimum is reached
- One of the optional termination conditions [***section 5.7***] is satisfied

The other buttons load/hide various editors or tables. These are listed below:

Icon	Chapter	Function
	***3***	Show/hide the variable editor Allows user to select variables & [optional] limits for those variables
	***4***	Show/hide the merit function editor Allows user to create/edit/review the smart merit function
	***5***	Load the optimisation options dialog to allow user to choose optimisation ending methods, optimisation strategy, backup frequency etc
	***4.2.1***	Table to show rays used by the smart merit function. [these rays are automatically maintained defined]
	***3.2.1.2***	Change table for current variables or for selected ranges of parameters.
	***3.2.1***	Table showing efficiencies for current variables or selected ranges of parameters. Useful in selecting variables.
	***3.2.2***	Table showing impact of asphberisation in optimisation. Useful in selecting surfaces for aspherisation.
	***6.1***	List of messages generated during optimisation.

3 Variables & Limits

- *The variable editor - defining variables*
- *The variable editor – defining hard limits which cannot be violated*
- *Choosing variables for optimisation and selecting surfaces for aspherisation*

"The most overlooked advantage of owning a computer is that if they foul up there's no law against whacking them around a bit."

- Eric Porterfield

One essential for optimisation is the ability to select variables; those parameters which will be automatically altered to improve the system performance.

In this chapter we will discuss how the user can define variables and also specify upper & lower limits for those variables. **Such 'hard' limits cannot be violated in optimisation**. The program will 'freeze' such variables at the specified limit, for a number of cycles.

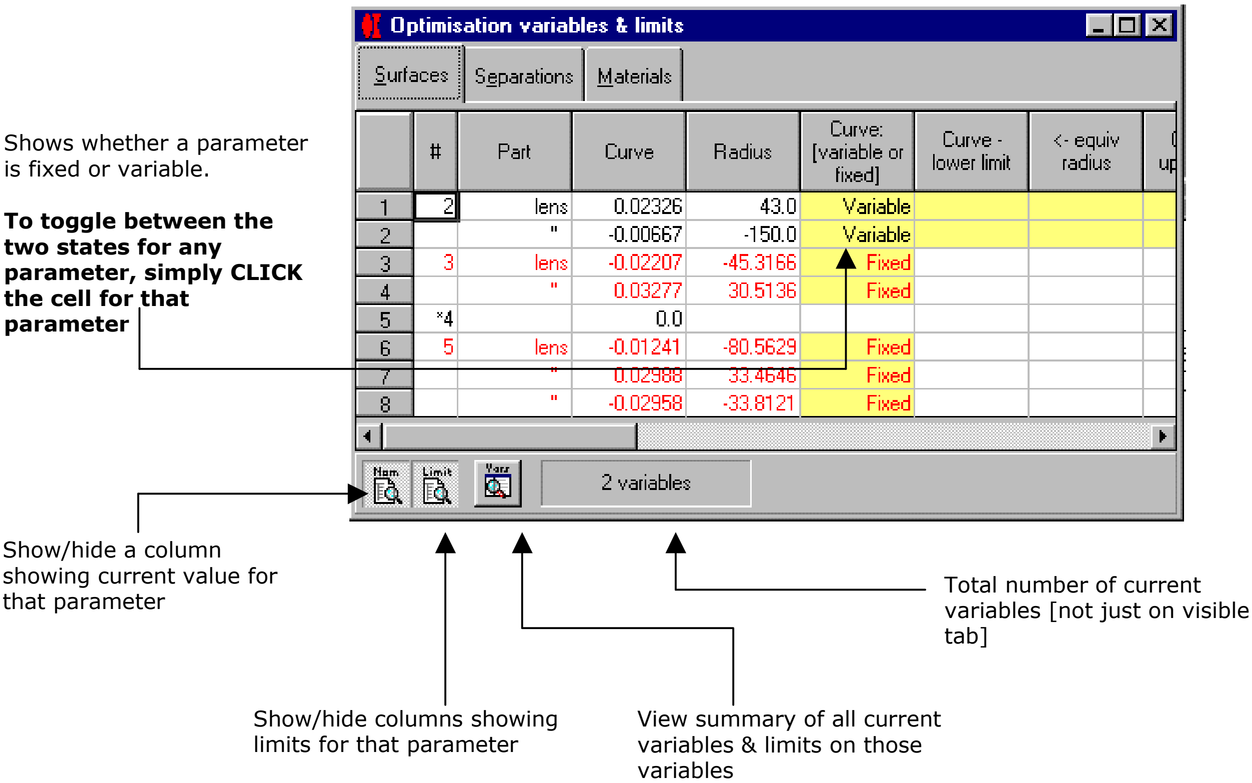
Although not technically a variable, we also discuss control of edge separations in any space [see ***section 3.1.2.1*** & ***section 5.2*** (default values)]

We will also discuss some new tools available within WinLens. These tools are designed to help the user choose variables for optimisation [***section 3.2.1***] and to choose surfaces for aspherisation [***section 3.2.2***].

3.1 The variable editor

The variable editor is a tabbed form[1], with one tab for each of the major groups of variables [surfaces, separations & materials]. In each tab, there is a spreadsheet listing these parameters component by component or surface by surface.

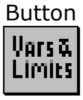
Below we show the editor as it would be seen on first loading, with a triplet system.



Setting variables is easy; as shown in the graphic, simply click in the cell relating to the parameter of interest.

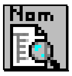
---

[1] To load the variable editor:

Menu	Select the 'Variables' option from the optimise menu
Button 	Click the merit function icon in the optimisation tab

Once a parameter has been made variable, the user may wish to set limits for the allowable range of values for that parameter. These limits will be used in the optimisation and cannot be violated!

Press buttons to show:



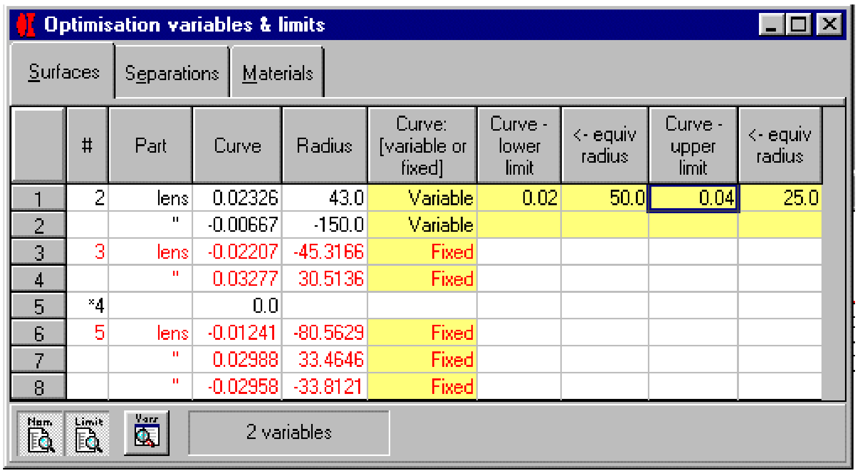
parameter values

limits on variables

list of current variables

Setting limits is almost as easy as selecting variables; the user only has to enter a number rather than just click a cell. Note that zero is acceptable as a limit.

To do this he must show the columns containing the limits. It is probably also a good idea to show the columns with the current values for the parameters.



	#	Part	Curve	Radius	Curve: [variable or fixed]	Curve - lower limit	<- equiv radius	Curve - upper limit	<- equiv radius
1	2	lens	0.02326	43.0	Variable	0.02	50.0	0.04	25.0
2		"	-0.00667	-150.0	Variable				
3	3	lens	-0.02207	-45.3166	Fixed				
4		"	0.03277	30.5136	Fixed				
5	*4		0.0						
6	5	lens	-0.01241	-80.5629	Fixed				
7		"	0.02988	33.4646	Fixed				
8		"	-0.02958	-33.8121	Fixed				

In the example above, for the same triplet, we show the surface tab[2]. Two radii have been defined as variables and can therefore have limits [if so desired]. Parameters that are not variable cannot have limits[3].

The first surface has a lower curvature limit of .02 [equivalent radius limit of 50mm] and an upper curvature limit of .04 [equivalent radius of 25mm]. If controlled in optimisation it can float between these two values without penalty. If however, it exceeds a limit, then the merit function contribution grows.

To define an upper or lower limit, simply place the cursor in the desired cell and enter the new value.

The surface tab is a special case – for technical reasons, WinLens has to control the curvature, not the radius. However most users are happier entering radius values, therefore WinLens shows the curvature value & limits, along with the radii equivalents.

In the following sections we will briefly cover the variable types and special features [if any] on each tab.

---

[2] Since all elements in this triplet have spherical surfaces, columns referring to aspheric variables & limits are automatically hidden.
[3] There is an exception to this rule for edge separations [***section 3.1.2.1***]

3.1.1 Surface variables

Within the restrictions defined in the notes below, the radii on all surfaces may be made variable.

Variable	Notes
Radius	Surfaces in blocks, LINOS components, and the stop [if surrounded by air] cannot be made variable. Limits are in terms of curvature

3.1.1.1 Surfaces with aspherics

When a surface is defined as a conic or an aspheric then the extra parameters defining those surface types are available as variables:

Variable	Notes
Conic Constant	
A4 [B4 or C4]	Coefficient offered depends upon global choice of coefficient type [as described below]
A6 [B6 or C6]	"
A8 [B8 or C8]	"
A10 [B10 or C10]	"
A12 [B12 or C12]	"

The coefficient displayed depends upon the current aspheric definition[4] in use. The limits [if any] will be appropriate for a particular definition.

NB Therefore if the user elects to change asphere definition**, limits on aspheres should be checked carefully**, as they could be wrong by dramatic amounts!

[4] Aspheres may be defined in terms of the conventional power series, a normalised power series, or a Zernike polynomial definition. This definition applies to the system as a whole.

3.1.2 Separation variables

NB Defocus variables & limits may be set in this tab.

The defocus values are shown in the last row.

See ***section 3.1.2.3***

This applies to separations within and between components. Note that separations between components can vary with zoom, but the axial separation inside an element is zoom independent.

Variable	Notes
Separation	Does not apply to LINOS components.

Zoom gaps are a special case and are discussed below.

3.1.2.1 Edge separation limits

In any design it is crucial to make sure that the component is physically viable. Thus, in the triplet example, no axial separations may be negative! This may easily be controlled by setting a lower limit on all variable separations.

However, the axial separation is not the only quantity that may go non-physical during optimisation. It is possible that surfaces may approach too closely or curve too much so that they 'cross' leaving a negative edge thickness.

The edge thickness of a space is the result of several different design parameters and also ray height and is therefore not directly handled in this editor. I.e. it is not possible to set hard limits on an edge thickness.

However 'soft' limits [extra entries in the merit function] can be added to control these values. Being 'soft' edge thickness violations may occur, and can only be controlled by reducing the tolerance on that defect.

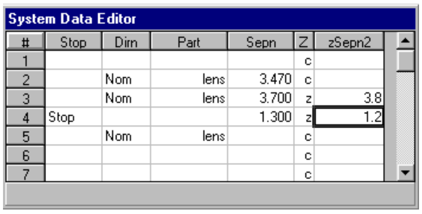
The user may add edge thickness defects manually, but automatic edge control is also available [***section 4.2.3.1***] in the optimisation.

Further notes on using automatic edge control defects are given in ***section 8.3***

Default values for the minimum edge thickness may be defined [***section 5.2***]. However that user may also enter custom values for the minimum edge limits as targets in the merit function editor.

3.1.2.2 Zoom gaps

Zoom gaps are special case. In a zoom system, the user may define some gaps [between components] as zoomable. Such gaps will have different separations in different zooms.

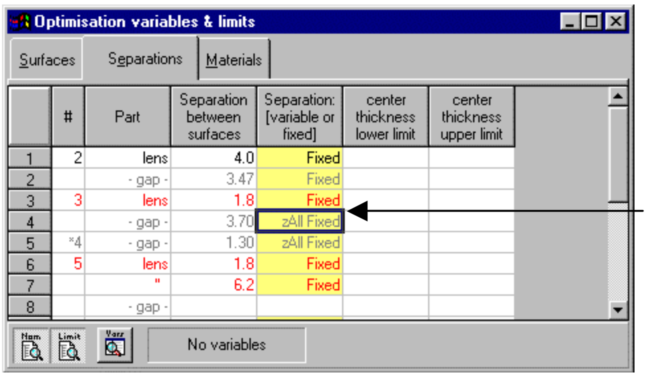

System Data Editor

#	Stop	Dim	Part	Sepn	Z	zSepn2
1					c	
2		Nom	lens	3.470	c	
3		Nom	lens	3.700	z	3.8
4	Stop			1.300	z	1.2
5		Nom	lens		c	
6					c	
7					c	

An example of a triplet with zoom-able spaces round the stop is shown here.

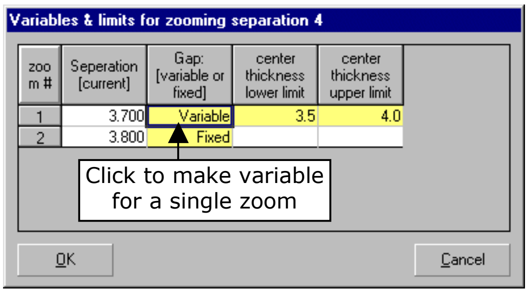
The user may wish to optimise the zoom-able separation values in none, some or all zooms.

The variable editor allows the user to make these choices. In such a case, the variable editor will show some differences.



Variable editor showing separations. Here all are fixed including the zooming gaps [denoted by 'zAllFixed]

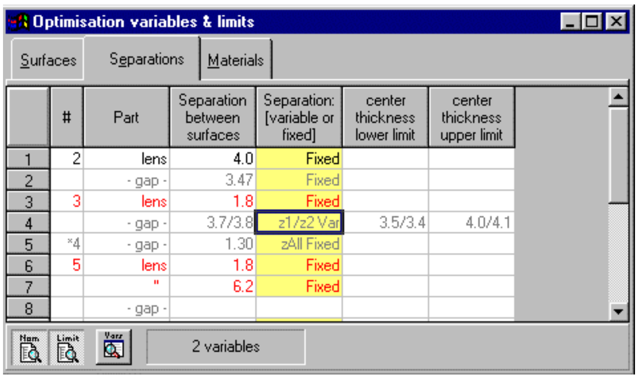
Click here to launch dialog to edit variables/limits on a single zooming gaps



Dialog to edit variables/limits on a **single** zooming gap.

[here separation 3 is variable in zoom 1 and is intended to lie within 3.5 & 4.0mm]

Once a separation has been made variable in a specific zoom the limits can be edited.



Editor showing system with zooming gap variables.

Here separation 4 is zoom-able & now has variables. In fact this separation has been made variable in 2 zooms.

If a zoom-able gap is variables in one or more zooms, then the editor displays data for those particular zooms [here zoom 1 & 2]

3.1.2.3 Defocus

For convenience, the defocus variables are shown in the same tab as the separations. The values are shown in the last row of the spreadsheet.

If the system has one zoom only, then both variable & limits may be directly entered into the spreadsheet. To set/clear variable defocus, then simply click on the defocus variable cell. To enter limits, simply type the value in the appropriate limit cell – of course, limits may only be specified when defocus is variable.

If the system has more than one zoom, then the same mechanism is used as discussed in the previous section on zoom gaps. Click on the defocus variable cell. WinLens will then launch the dialog in which the user can set/clear variables for any zoom, and then set/clear limits for any zoom with variable defocus.

3.1.3 Material variables

It is possible to make variables of the materials inside the system. This applies to:

- materials inside components
- materials between components

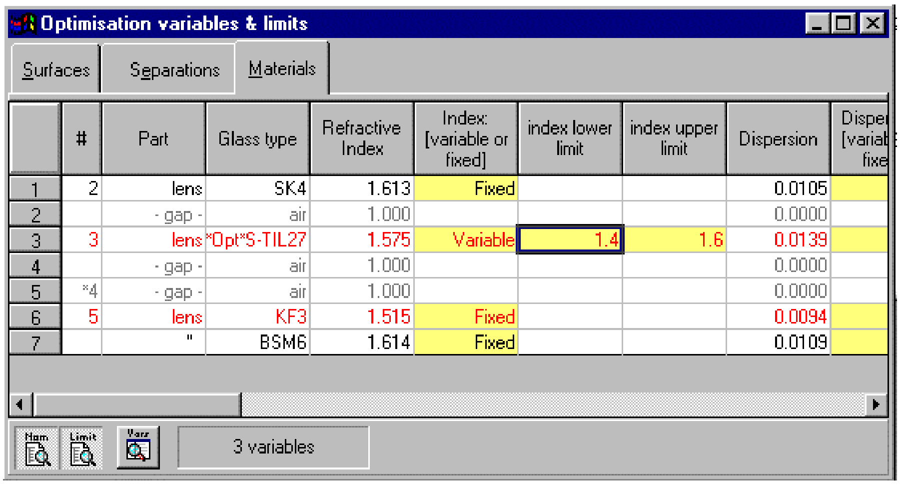
It does not apply to air spaces within or between components.

Variable	Notes
Index	Does not apply to materials inside LINOS components
Dispersion	Does not apply to materials inside LINOS components

Without limits materials would immediately run toward totally unphysical solutions with very high indices and very low dispersions. It is therefore critical to keep optimised materials within regions where real glasses exist.

It is possible to specify limits on those materials. Such limits on index and dispersion variables will simply keep glasses within specified rectangular regions on the glass map.

However if automatic glass control is selected [***section 4.2.3.2***] then glasses will be kept within the real glass region of the glass map. One or both methods may be used.



Note that when a glass is made variable [either by index or dispersion], the glass name is prefixed with '*Opt*', to indicate that the material is based upon the named glass, but may differ in actual values.

Thus, in the example above, the glass S-TIL27 was used in the original design. Of course no changes are made until the optimisation begins.

3.2 On the choice of variables

In the previous section we discussed the mechanics of defining variables. In this section we discuss how to select 'good' variables. We also discuss a guide for selecting surfaces for asphérisation.

3.2.1 Variable efficiencies & change table

Both of these tools require that the merit function [***Ch 4***] be defined.

The choice of variables for optimisation is obviously very important. Equally obvious is that fact that the writers of optimisation software normally ignore this problem.

However in WinLens, we do offer help in making these choices. A word of caution; the tools offered are not magic!

The user may view two tables: the change table and the variable efficiency table. Both are updated on demand only [i.e. unlike most other tables which update after a change in the lens design]. The user may elect to view results for either:

- Current variables
- Selected ranges of parameters [e.g. all radii and/or all separations etc, etc].

The selection applies to both tables.

3.2.1.1 Variable efficiencies

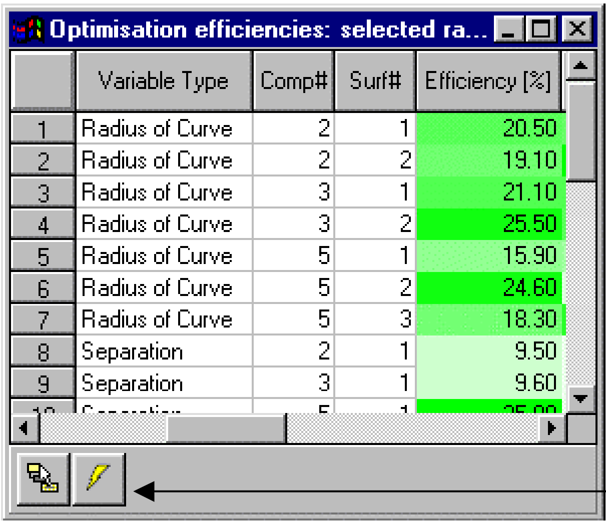
The concept of variable efficiencies is quite straightforward. The defect vector, f, is a vector in defect space. It describes the difference between the defects of the current system and the target values of those same defects. Each variable has an associated vector, $\underline{a}_i$, which shows the change in defect as a result of a small change in that variable.

If $\underline{a}_i$ is parallel/anti-parallel to the defect vector, then it is going to be very useful or efficient in moving the solution toward the desired target values.

If however $\underline{a}_i$ is perpendicular to $\underline{f}$, then changing that variable is not going to reduce the merit function; it is not therefore efficient.

To quantify this we define the variable efficiency. The variable efficiency is the cosine of the angle between the defect vector and the defect vector, $\underline{a}_i$. It is displayed as a percentage.

It is more sensible to choose variables with a high efficiency, as they should, on average, lead to larger or more rapid drops in merit function



	Variable Type	Comp#	Surf#	Efficiency [%]
1	Radius of Curve	2	1	20.50
2	Radius of Curve	2	2	19.10
3	Radius of Curve	3	1	21.10
4	Radius of Curve	3	2	25.50
5	Radius of Curve	5	1	15.90
6	Radius of Curve	5	2	24.60
7	Radius of Curve	5	3	18.30
8	Separation	2	1	9.50
9	Separation	3	1	9.60

Colour coding of parameters: greener parameters are more efficient.

To sort the parameters: click on a column header.

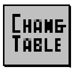
Setup options for analysis & display [i.e. choose to review current variables or selected ranges of parameters]

Force update of variable efficiency table

Note that the variable efficiencies will almost certainly change during optimisation, so those variables, which were useful at one point, may not be useful later on.

Note also that ranges of aspheric coefficients are not analysed in this table, since they are handled in the asphérisation guide [***section 3.2.2***]

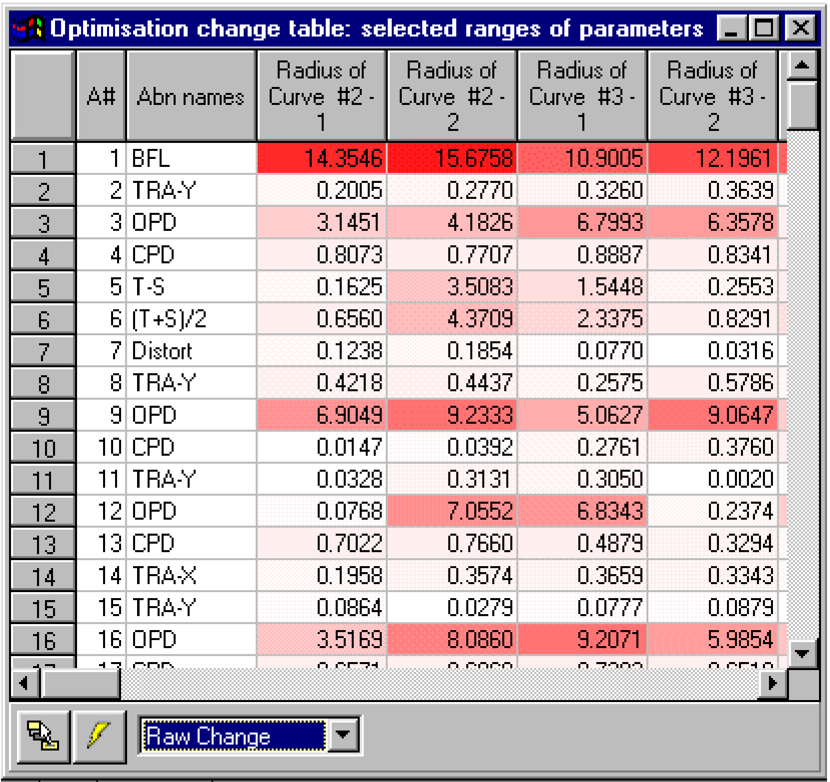
3.2.1.2 Change tables



The change table is a display of the changes in each defect as a result of a small increment to each parameter. This is often known as the sensitivity table.

Colour coding of sensitivities:
redder values have larger changes.

To sort the sensitivities:
click on a column header.



	A#	Abn names	Radius of Curve #2 - 1	Radius of Curve #2 - 2	Radius of Curve #3 - 1	Radius of Curve #3 - 2
1	1	BFL	14.3546	15.6758	10.9005	12.1961
2	2	TRA-Y	0.2005	0.2770	0.3260	0.3639
3	3	OPD	3.1451	4.1826	6.7993	6.3578
4	4	CPD	0.8073	0.7707	0.8887	0.8341
5	5	T-S	0.1625	3.5083	1.5448	0.2553
6	6	(T+S)/2	0.6560	4.3709	2.3375	0.8291
7	7	Distort	0.1238	0.1854	0.0770	0.0316
8	8	TRA-Y	0.4218	0.4437	0.2575	0.5786
9	9	OPD	6.9049	9.2333	5.0627	9.0647
10	10	CPD	0.0147	0.0392	0.2761	0.3760
11	11	TRA-Y	0.0328	0.3131	0.3050	0.0020
12	12	OPD	0.0768	7.0552	6.8343	0.2374
13	13	CPD	0.7022	0.7660	0.4879	0.3294
14	14	TRA-X	0.1958	0.3574	0.3659	0.3343
15	15	TRA-Y	0.0864	0.0279	0.0777	0.0879
16	16	OPD	3.5169	8.0860	9.2071	5.9854

Setup options for analysis & display, i.e. choose to review:
- current variables
- selected ranges of parameters

Force update of change table & variable efficiency analysis table

Choice of display, either:
- Raw change data
- Sensitivities
- Normalised sensitivities

The change table [shown above] provides a more detailed view of the sensitivity of each defect to each parameter, while the variable efficiency table [previous section] provides an overview.

Three views of the data are available:

- Raw changes in i^{th} defect induced by the j^{th} parameter change [y_I - y_{Ij}*].
- Sensitivities: f_{Ij} (= [y_I - y_{Ij}*]/tol_i) [where tol_i is the tolerance on the i^{th} defect]
- Column normalised sensitivities: [f_{Ij} /sqr(sum(f_{Ij}^2 all i)] used in optimisation

For the raw changes, the differences for all defects are shown. But for both type of sensitivity, only data for active defects with non-zero tolerances are shown.

Note: because both the variable efficiency table and this table share the same basic data, forcing an update on one will cause an update on the other.

Note: during optimisation sensitivities are calculated for every variable at the beginning of each cycle. Therefore if loaded [and the option selected] these forms will be updated at the beginning of each cycle, to show the data for the current variables. At the end of optimisation these are not updated automatically, as this would require an extra sensitivity analysis. However the user may elect to force a manual update.

3.2.2 Asphérisation guide

Many systems do not require aspheric surfaces. However for many other systems the use of one or more asphericals can offer significant advantages in performance, weight or some other benefit.

Sometimes the choice of surface for aspherisation is obvious, but this is not always true. Moreover, the necessary degree of the aspheric is rarely obvious.

In WinLens, we offer a tool to help with these choices: the aspherisation guide.

The user may choose which orders of aspheric coefficients are to be analysed, e.g. he may elect to analyse the impact of varying conic constant, A4 & A8 coefficients

The aspherisation guide, when an analysis is requested, works through each surface in turn. Changes are made to that surface alone and results calculated; the design parameters are then returned to the nominal values.

For each surface, we find

- Variable efficiency of each chosen order of aspheric coefficient
- Reduction of merit function after 1 optimisation cycle, varying only yhe chosen coefficients on that surface.

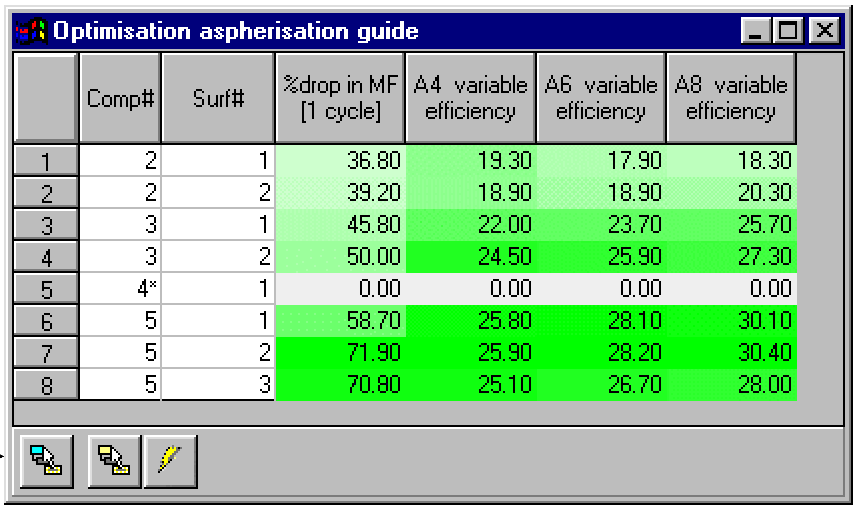
A typical example is shown below:

Colour coding of drop in merit function:
greener values shows larger reduction.

Colour coding of variable efficiency:
greener values shows more efficient parameters.

To sort the values:
click on a column header.

Load dialog to choose which orders of aspheric will be analysed



	Comp#	Surf#	%drop in MF [1 cycle]	A4 variable efficiency	A6 variable efficiency	A8 variable efficiency
1	2	1	36.80	19.30	17.90	18.30
2	2	2	39.20	18.90	18.90	20.30
3	3	1	45.80	22.00	23.70	25.70
4	3	2	50.00	24.50	25.90	27.30
5	4*	1	0.00	0.00	0.00	0.00
6	5	1	58.70	25.80	28.10	30.10
7	5	2	71.90	25.90	28.20	30.40
8	5	3	70.80	25.10	26.70	28.00

Load dialog to make global choice of aspheric definition for entire system

Force update of asphérisation guide

In this case we see that aspherising the final surface gives a much larger reduction in merit function that aspherisation of the first surface. This is therefore probably a better choice for an aspheric surface.

4 Merit function

- *Definition of a merit function*
- *Smart features of the WinLens merit function*
- *Guide to the merit function editor*
- *Setup wizard for default merit functions*
- *Editing the merit function*
- *Active/inactive defects*
- *Sorting defects by contribution*

"All sorts of computer errors are now turning up. You'd be surprised to know the number of doctors who claim they are treating pregnant men."

- Isaac Asimov

4.1 What is a merit function

The performance of an optical system can be measured by many different quantities or defects. Some of these, such as focal length, will often have target values, while others should be as near zero as possible.

For the purposes of optimisation, we need to combine all these measures into a single number, by what is known as a merit function. Within WinLens the merit function, f, is defined as:

$$f(x_i,..x_n)=\sum_{j=1}^{m}\left(\frac{y_j(x_i,..x_n)-t_j}{tol_j}\right)^2$$

where:

x_i	value of the ith system parameter [e.g. radius, separation, etc]
y_j	value of the jth defect
t_j	target value of the jth defect [often zero]
tol_j	acceptable tolerance on the defect

Once the defect type has been defined, the user may specify the desired target value and also the acceptable tolerance on the difference between the actual and desired defects. Note that the tolerance is a physically meaningful quantity!

For example, suppose the focal length is important. This will then be one of the defects composing the merit function. The user may require a focal length of 250mm; this becomes the target. Often the focal length is controlled to within+/-1%, so in this case the tolerance is +/-2.5[mm].

Similarly if the performance requirements call for a spot diameter of less than 0.1mm, then the targets on transverse rays aberrations will be zero and the tolerance will be +/- 0.05[mm] or smaller.

As the system 'improves', the defects get closer to the target values and the merit function gets smaller. The aim of optimisation is to find a minimum value for the merit function.

Clearly it is important that the merit function adequately models/monitors the performance of the system. Thus a merit function entirely composed of paraxial quantities would provide no control over image quality! The optimisation would quickly yield a minima, but the actual lens would be useless.

Obviously only the lens designer knows what is appropriate/necessary for his lens. Therefore the WinLens merit function is very flexible; any of the available defects can be added in any order [***section 4.5***].

However, to provide a good starting point, either of two setup wizards will help create a merit function [***section 4.4***], which can then be edited [***section 4.6***].

The merit function in WinLens is smart [***section 4.2***] and is supported by a sophisticated merit function editor [***section 4.3***].

Although the merit function is always saved with the lens data, it is also possible to save a merit function separately as a .MFN file. These can be used by other systems. A preview facility is available to allow the user to load part or all of the disk based merit function into the 'clipboard' for subsequent paste operations. [***section 4.6.3***]

4.2 The WinLens smart merit function

The WinLens optimisation routines have been written with the aim of keeping time sent on maintaining the merit function to a minimum. A number of special features have been incorporated.

4.2.1 Automatic ray set definition

Many defects are real ray based, i.e. require particular rays to be traced. For example, a distortion or astigmatism aberration requires a chief ray, while a transverse ray aberration requires both a chief ray and a skew ray at the appropriate aperture co-ordinate.

In the old days the user had to create a ray set and then force the defects to reference the rays by number. This, of course, was a very error prone procedure – frequently defects would reference the wrong ray!

This is no problem in WinLens. The set of rays required by the defects in the current merit function is automatically generated. The rays can be reviewed by loading the optimisation ray table.

4.2.2 Self maintenance

Some defects refer to specific surfaces. Other types of defect use earlier defects in the merit function. A simple merit function with these types of defects could be severely disrupted by adding/removing/reversing components or cutting/pasting other defects.

The WinLens smart merit function knows about these links and keeps them correct under editing.

If such self-correction is impossible, e.g. by removal of a component, which is referred to by a defect, then a clear warning is given. The user could then choose to remove the offending defect or assign it to another surface.

4.2.3 Automatic defects [optional]

Defects added by the user by drag drop form the defect list or by pasting or by use of the setup wizard, are all known as 'user' defects. It is also possible to get the editor to automatically add extra defects at the end of the merit function [***section 4.2.3***] to control edge thickness & glass variables.

4.2.3.1 Control edge thickness

Edge defects: applies to all spaces. Aim to keep edge thickness greater than a minimum value [default minimum values may be defined as part of the optimisation options – ***section 0***]. Used to make sure the system is physically possible and that the system does not have negative edge separations! For zooming gaps, one such defects is added/zoom.

Note: if you wish to keep an edge thickness less than another value you can add thickness defects manually.

4.2.3.2 Control glass variables

Glass defects: applies to any variable glasses. These aim to keep the glasses within the region [on the N_dV_d map] of real glasses [The function governing the contribution to the merit function is shown in ***section 5.4***].

Note that if a glass is variable and you have defined limits for the index and/or dispersion, it will be controlled by a variable defect and a glass defect. This may be useful if you wish to restrict the glass to a particular rectangle within the real glass region.

4.3 A guide to the Merit Function editor

To support the WinLens[5] merit function, we have provided a sophisticated merit function editor and a setup wizard. The editor therefore has quite a few different options. An exploded view of the editor is shown below.

List of available 'defects' arranged in 'families' ***Section 4.3.2***

Spreadsheet showing merit function ***Section 4.3.1***

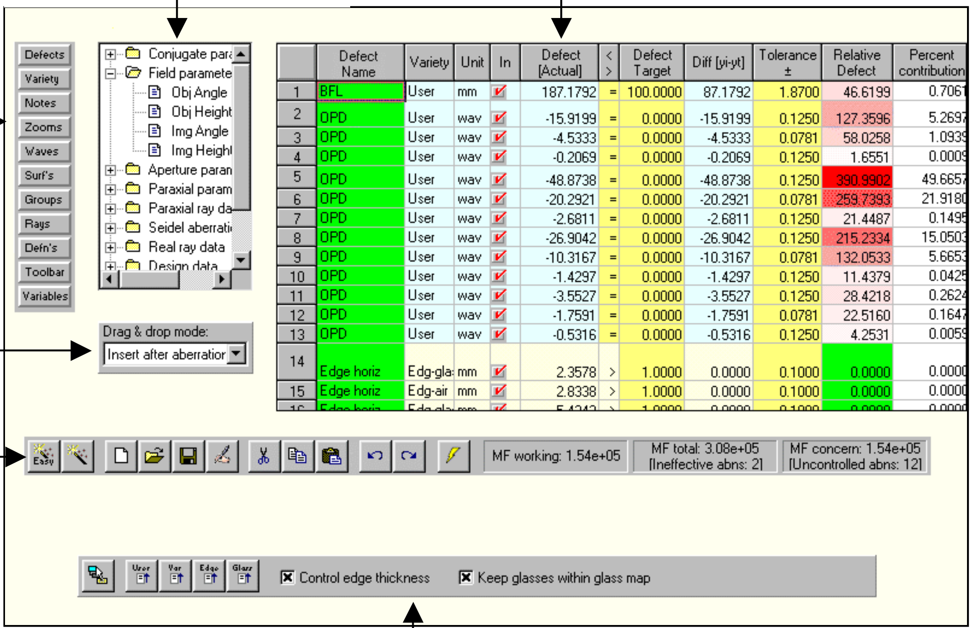
Buttons to show or hide various columns in the spreadsheet ***Section 4.3.4***

Insert mode when dragging defects from tree list to spreadsheet ***Section 4.3.3***

Editing toolbar buttons:

- Quick setup wizard
- Full setup wizard
- File IO
- Editing options
- Undo/Redo
- Update

Section 4.3.5
Also merit functions

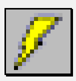


Options toolbar for automatic aberrations:

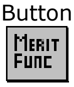
- Buttons to jump to first defect of that variety
- Check box to include/exclude automatic defects of that variety

Section 4.3.6

Two particularly important functions are:

	Load the quick setup wizard to create a default merit function ***Section 4.4***
	Update the current merit function [includes validating all defects and generating the associated ray set]

[5] To load the merit function editor either:

Menu	Select the 'Merit function' option from the 'Optimise' menu
Button 	Click the merit function icon in the optimisation tab

4.3.1 Spreadsheet

The merit function is displayed in a spreadsheet within the merit function editor. Each defect is defined within a single row. Each column either:

- shows a particular parameter relevant to that defect
- shows a calculation value for that defect [e.g. merit function contribution]

Not all defects use all parameters. Where a parameter is not relevant to a defect that cell will be left blank.

Not all columns will be visible at all times. Some may be deliberately hidden by the user [***section 4.3.4***], while others only appear during optimisation [such as the prediction of the defect value]. Working from left to right:

Header	Meaning	Further notes
Defect Name	Descriptive name of defect [defect]	***Section 4.3.2***
Variety	All defects are either: • user defined • automatic [variable, edge thickness, glass]	***Section 4.2.3***
Unit	Unit for defect e.g. mm, wave etc	
Notes	User can enter any comment or note here	
In	Defects are active [checked] or inactive	***Section 4.7***
Zm#	Zoom: most defects are for a single zoom	
Wv#	Wave: some defects are for a particular wave	
Srf#	Surface: some defects relate to a specific surface	
Srf#	Surface2: some defects also specify a second surface	
Grp#	Group/module: a few defects specify this	***Page 69***
Ry#	Ray: [number in table of optimisation rays]	
FieldY	Relative field co-ordinate of ray	
ApX	Relative X aperture co-ordinate of ray	
ApY	Relative Y aperture co-ordinate of ray	
Combined defect defn	Definition of a defect taken from the 'combined' family of defects	" & ***Section 4.3.2***
< >	'direction' of defect. Three options determine calculation of contribution of defect to Merit Fn =: contrib = (actual-target)/tolerance [**default**] >: contrib as default IF target > actual, otherwise 0 <: contrib as default IF target < actual, otherwise 0	
Defect [Old]	[*Seen in optimisation only*]. Value of this defect at end of previous cycle	
Defect [Pred]	[*Seen in optimisation only*]. Value of this defect predicted using last solution and differentials	
Defect [Actual]	Actual value of this defect when last evaluated	
Defect Target	User defined target value for this defect	***Section 4.1***
Diff [yi-yt]	Difference between actual & target value of this defect	
Tolerance	User defined +/- tolerance on the defect	***Section 4.1***
Relative Defect	(actual-target)/tolerance for each defect. Merit Fn value is the sum of squares of these values	***Section 4.1***
Percent Contribution	[Relative defect]^2 as a % of the working Merit Fn This is only given for effective defects	
Non Linearity	[*Seen in optimisation only*]. Measure of the non linearity between the actual and the predicted defect	

To help understand the spreadsheet liberal use is made of colour. The spreadsheet may also be edited directly to change single parameters for individual or groups of defects.

Cut/copy/paste operations for whole defects [one or more rows] are discussed elsewhere [***section 4.6***]

4.3.1.1 Colour coding

There are three separate uses for colour in the spreadsheet. These are:

- Distinguish editable/non editable cells. Editable cells [such as the tolerance & target columns] have a yellow background.
- Show valid/invalid defects: Defect name column. If the Defect definition is valid the background is green. If invalid then red. If not tested then dark grey.
- Show relative magnitude of merit function contribution. Merit Fn Contrib column. Defects within tolerance have a clear green background. Defects outside of tolerance have a red->white background. The largest contribution to the merit function has a clear red background. All others have some shade of red/pink which is determined by the relative contribution, i.e. the redder the background the worse the defect.

4.3.1.2 Editing individual parameters

Some values in the spreadsheet may be edited directly, such as the target and tolerance values for each defect. Simply click on the cell of interest and type in the required value.

For other cells, e.g. Surface #, Zoom # etc, there is a limited range of possible values. A popup dialog will display these values. Simply click the required option.

4.3.1.3 Editing groups of parameters

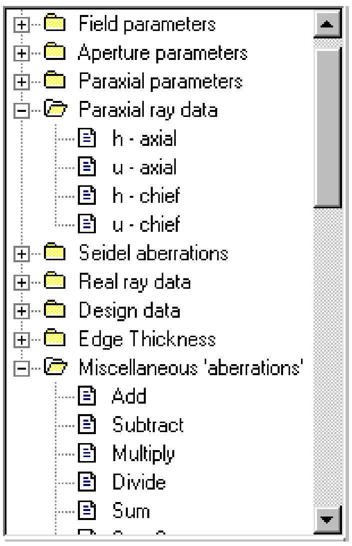
Often a user will want to simultaneously set the values for a single parameter for a range of defects. For example, he may want to change the Y relative aperture from 0.7 for a number of TRA defects, or even change some TRA's to OPD's.

Simply select the block of cells in the appropriate column. A dialog will appear allowing the user to specify the new value for that parameter. [NB if you change the actual defect type, all other parameters will be unchanged]

If that parameter does not have a fixed range of values, then the user may elect to choose a new value or to scale the old value. Furthermore, he may elect to apply the change to all defects in the range, or only to specific types of defect within the range [a fuller discussion is found in ***section 4.6.2***].

4.3.2 Defect list

The merit function is composed of one or more defects taken from an extensive selection. Rather than having to remember cryptic codes, WinLens shows all the available defects, sorted by families, in a tree structure.



The user simply drags and drops from the list to the spreadsheet, when he wants to add a defect to the merit function.

Here we see most of the families, two of which have been expanded to show the individual defects.

Notes:

- Family – Real ray data: include defects that measure performance in image space and others that give ray coords at a surface
- Family - Design data: accesses parameters defining the system
- Family – Miscellaneous: set of defects that manipulate earlier defects in merit function

4.3.3 Insert mode

When a defect is dragged from the list to the spreadsheet, it is inserted at the 'drop point'. Two options are available

- Insert after defect clicked by mouse
- Replace defect clicked by mouse

Select the desired option from the drop down list which lives under the list of available defects.

4.3.4 View options

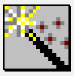
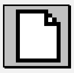
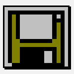
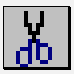
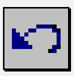
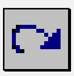
It is easy to show or hide different parts of the merit function editor. Arranged down the left hand side of the editor are a number of icons. Their functions are defined in the table below:

Icon	**Function**
Abn's	Show/hide tree list of available defects, arranged in families
Variety	Show/hide notes column
Zooms	Show/hide column in spreadsheet which display zoom information
Waves	Show/hide column in spreadsheet which display wave information
Surf's	Show/hide columns in spreadsheet which display surface information
Groups	Show/hide column which displays group/module information [only used by group/module defects: see ***page 69***]
Rays	Show/hide ray related columns in spreadsheet which display: • Ray # • Relative field co-ordinate • Relative aperture co-ordinates
Defn's	Show/hide Combination Defn. column in spreadsheet. This shows definition of one of the 'combination' defects, all of which relate to earlier defects in the spreadsheet.
Toolbar	Show/hide options toolbar
Variables	Show/hide NON-EDITABLE list of current variables. To actually set/clear variables use the variable editor [***section 3.1***]

Although not essential, these help the user see what is relevant at any one time.

4.3.5 Setup/editing toolbar

This toolbar provides many useful functions, which are described below.

Icon	Function
	Quick setup wizard: creates a default merit function [***section 4.4***]
	Advanced setup wizard: creates a more customised merit function, according to users preferences. This remembers last choice and is especially flexible for zoom systems.
	New merit function: clears existing merit function
	Open merit function file: allows user to load a merit function file [*.MFN]. This is displayed in a preview table. User can choose to load this directly, or to copy some or all of the merit fn to the clipboard.
	Allows user to save the current merit function to disk as a .MFN file. Note that the merit function is always saved in the lens file – so this is a separate file.
	Allows user to make a note attached to the merit function.
	Cut function. Enabled if user has selected one or more whole defects. Cuts selected defects from the merit function. The cut defects are placed on the clipboard for subsequent paste.
	Copy Function. Enabled if user has selected one or more whole defects. Copies selected defects to the clipboard for subsequent paste.
	Paste Function. Enabled if defects are on the clipboard. Inserted after current row. If several rows selected in spreadsheet then these will be replaced.
	Undo Function. Multi level undo
	Redo Function. Multi level redo
	Update function. When pressed this • Re-validates merit function • Calculates defects • Updates spreadsheet

At the right hand side of the toolbar, WinLens Plus displays up to three versions of the merit function value, along with other information.

It may seem very strange to display three different merit functions, so we will define these carefully. Working from left to right:

- MF working: $\sum$(contribution from all effective defects)^2
- MF total: $\sum$(contribution from all defect's)^2
- MF concern: $\sum$(contribution from all effective defects outside of tolerance)^2

From these definitions, it is clear that MF total ≥ MF working ≥ MF concern

MF working is that actually used in the optimisation.

The other two [MF total & MF concern] are displayed for extra information.

It is possible to terminate the optimisation when MF concern=0 [i.e. all effective defects are within tolerance] [to select this option, see ***section 5.7***]. The idea behind this is that once all defects are controlled, there is no need to go any further.

The number of ineffective defects [zero tolerance or marked as inactive], is shown in the same cell as the MF total. If there are no ineffective defects, then MF total is the same as MF working, and is not shown.

Finally, the number of effective defects, which are outside limits [known as uncontrolled defects], is shown in the same cell as the MF concern.

4.3.6 Options toolbar

This toolbar contains a number of icons and three checkboxes. These are basically related to the display/use of the automatic defects [***section 4.2.3***]

Icon	Function
	Shows summary of optimisation options and allows user to load optimisation options dialog
	Scrolls spreadsheet so that first 'user defined' defect is visible at the top of the spreadsheet.
	Scrolls spreadsheet so that first 'variable control' defect is visible at the top of the spreadsheet.
	Scrolls spreadsheet so that first 'edge thickness control' defect is visible at the top of the spreadsheet.
	Scrolls spreadsheet so that first 'glass control' defect is visible at the top of the spreadsheet.

Check Box Captions	Function
'Control edge thickness'	When checked, the editor automatically adds & maintains defects to keep edge thickness above minimum limits. [***section 4.2.3.1***]
'Keep glasses within glass map'	When checked, the editor automatically adds & maintains defects to keep glass variables within the 'real' glass region of the glass map. [***section 4.2.3.2***]

4.4 Setup wizards & the merit function

Although a merit function can be assembled by hand [***section 4.5***], two setup wizards are also available to simplify the task.

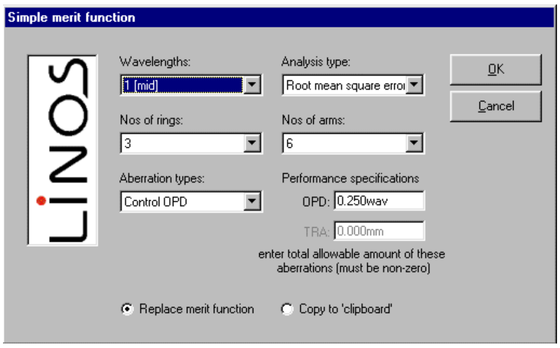
The quick setup wizard can generate a simple balanced default merit function with one mouse click [though several options, requiring a few more clicks, are available on the dialog!]. To obtain this simplicity of operation, several assumptions are made, but these are generally reasonable.

Alternatively, the full setup wizard can be used to create a custom merit function. This wizard remembers the last set of options used, and is particularly flexible when dealing zoom systems.

Whichever wizard is used, the user may choose:
- to replace any existing merit function with the new defects
- to place the new function on the clipboard for subsequent pasting operations

4.4.1 Quick setup wizard

The quick setup wizard is a single stage dialog.



The pre-selected options will generate a merit function, which aims to yield a diffraction limited system [OPD<0.25 waves], with controlled back focal length

The ray patterns generated by this wizard are covered in ***section 4.4.2.3*** below]

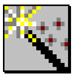
The quick wizard has some degree of flexibility, in that the user may choose:
- number of analysis wavelengths [1, 3 or 5]
- number of rings in ray pattern at each field point
- number of arms in ray pattern at each field point
- whether to control RMS or peak to valley aberration
- whether to control OPD, TRA or both

The user must also enter a maximum specification for the total amount of OPD and/or TRA. Thus if the user enters a specification for the TRA of 0.01mm, the tolerances on the TRA in the merit function will be +/- 0.005mm or smaller.

To achieve this simplicity, the program deliberately limits the options available:
- The ray patterns are based on vignetting of the current system
- All zooms are treated identically
- All fields are equally weighted
- Only the ring ray pattern is available

If the user wishes to create a merit function based upon different assumptions then he must use the advanced wizard.

4.4.2 Full setup wizard



The full setup wizard is a multi page dialog. It remembers the last set of options selected and will use these as the starting selection, unless the user clicks the 'default options' button on the first page. It is designed to allow for different settings for different groups of zooms.

Clicking the 'Next' button reveals the next appropriate page:

- The first few pages will only be seen for zoom system [***section 4.4.2.1***].
- The next set of pages will be shown for all systems [***section 4.4.2.2***].
- The next pages will depend upon the user choice of ray pattern [rings of rays [***section 4.4.2.3***] or fans of rays [***section 4.4.2.4***]];
- The final page allows the user to choose between overwriting the existing merit function, and putting the next merit function on the clipboard

4.4.2.1 Setup wizard: zoom pages

By default, separate sets of defects within the merit function are added for each zoom. Also, by default, the same settings and choices are used for each zoom. This may not be appropriate for your system, so you can:

- Specify which zooms are to be used [some can be ignored]
- Create groups of zooms [from those specified] to have same settings
- Select settings and choices for each group of zooms.

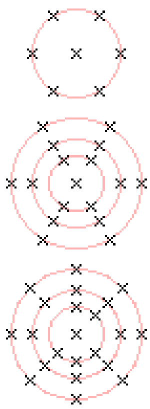
That final stage simply involves choosing one of the groups of zooms and working through the pages described in the next sub sections. At the end of this, the user is allowed to return and choose another group for the same procedure. This continues until the user is happy with the settings for all groups of zooms.

4.4.2.2 Setup wizard: standard pages

The user is shown a list of the current settings and choices. He may skip to the end at this point. Alternatively, clicking the next button will allow him to change any or all of these settings. In sequence, these are:

Choice	**Comment**
Paraxial targets	A number of simple paraxial constraints that can be included in the final merit function
Number of wavelengths	Choice of 1, 3 or 5 wavelengths [as defined in the system parameter editor]. Ray pattern will be duplicated for each wavelength
Field Weighting	Often axial performance is more critical than off axis. User may assign relative weights to the fields.
Vignetting	User may choose to have aperture rays that: • are limited by vignetting of the system [at time of use of setup wizard!] • are limited to a relative paraxial aperture of +/-1 for all fields [i.e. ignore vignetting]
Performance measure	Choose whether system is diffraction limited or not. If diffraction limited then specify acceptable amount. If not diffraction limited then user will have to enter performance requirements during ray pattern definition stage
Ray pattern	Performance of system has to be analysed by representative ray patterns [one from each object point]. The user may choose between: • Rings of rays [***section 4.4.2.3***] • Fans of rays [***section 4.4.2.4***] Pages for the two patterns be different until the final page [***section 4.4.2.5***]is reached.

4.4.2.3 Setup wizard: ray rings [option A]



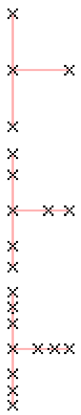
Ray Rings:
rings# & arm#

The system is probed by concentric rings of rays centred round a chief ray. This pattern is duplicated for each object point [as used in the standard graphs/tables]. This pattern is based upon a paper by Forbes[6].

The merit function contains terms [OPD and/or TRA] from each ray in the pattern.

Choice	Comment
Pattern definition	Choice of: • Number of rings • Number of arms • RMS or PeakToValley
Aberration type	Merit function to include OPD, TRA or both. If system is non-diffraction limited, then user will need to enter maximum acceptable OPD and/or TRA as appropriate.

4.4.2.4 Setup wizard: ray fans [option B]



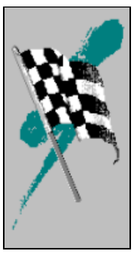
Ray Fans:
ray# [4, 7 or 10]

The system is probed by fans of rays centred round a chief ray. These fans lie in the meridian and sagittal planes defined by the chief ray and optical axis of the system This pattern is duplicated for each object point [as used in the standard graphs/tables]. This type of pattern has been in use for many years and was developed at Imperial College.

The merit function contains terms [OPD and/or TRA] from each ray in the pattern. It may also contain distortion and astigmatism aberrations from the chief ray.

Choice	Comment
Pattern definition	Choice of number of rays in the fans for each object point [4,7 or 10]
Aberration type & specification	User specifies general classes of defect to control. Wizard will determine appropriate specific defects to use. If system is non-diffraction limited, then user will need to enter maximum acceptable limits for these types.
Aperture weighting	User may specify how rays are weighted relative to their aperture location.

4.4.2.5 Setup wizard: finish



Once the final page has been reached and the user is happy with his choices he has one final choice to make before he presses the 'Finish' Button.

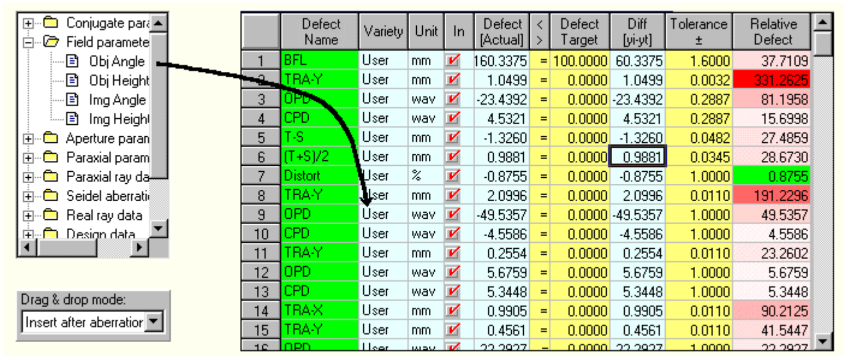
That choice is:
- new merit function completely replaces any existing merit function [default]
- new merit function is placed upon the clipboard. The old merit function [if any] is left undisturbed. Once the wizard has cleared, then the user may insert the new function into the existing function.

When the user presses the 'Finish' button, the wizard will disappear and the new merit function be generated. Please note that the user may be asked some further questions during this stage.

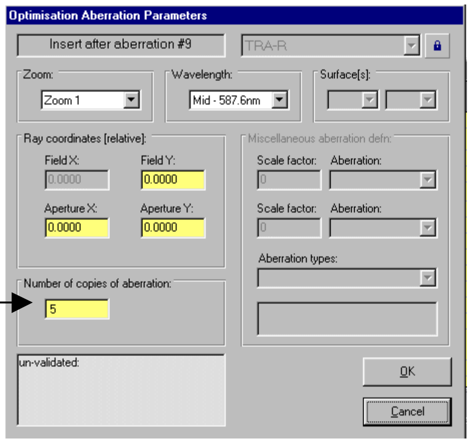
[6] GW Forbes 'Optical system assessment for design: numerical ray tracing in the Gaussian pupil, JOSA p1943, Nov 1988

4.5 *Adding extra terms to the merit function*

To add extra defects simply drag & drop from the defects list to the spreadsheet.



WinLens will then launch a dialog to enable the user to enter the parameters appropriate for that defect.



Number of copies of this aberration to be inserted

Part of the reason for doing this is to allow the user to create multiple copies of a defect in a single operation. Here, for example, 5 copies of the TRA-R defect are being added after existing defect # 9.

Once completed, WinLens will then validate the parameter choices and enter the new defect[s] into the spreadsheet.

At this point the user can edit individual parameters directly in the spreadsheet.

Alternatively by right clicking in the 'Defect Name' column, this dialog is loaded again. This is particularly useful for:

- editing the definition of a defect from the 'combination' family
- altering the defect type [click the lock icon at top left, and select from the defect list.

4.6 *Editing the merit function*

When optimisation programs were first written the user either had a very restricted merit function which was easy to use, or a very flexible function which was complex to set up and even more difficult to maintain, especially if it contained references to specific surfaces or to other defects.

WinLens Plus has a very flexible merit function, which is easy to create and maintain. A sophisticated editor provides the necessary environment. In this section we describe the various editing features.

4.6.1 Editing individual parameters

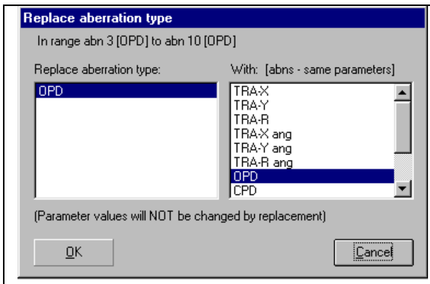
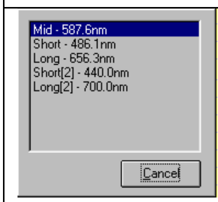
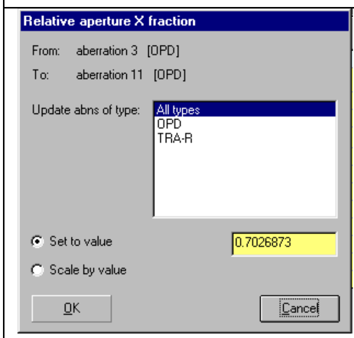
Some values in the spreadsheet may be edited directly, such as the target and tolerance values for each defect. Simply click on the cell of interest and type in the required value. For other cells, e.g. Surface #, Zoom # etc, there is a limited range of possible values. A popup dialog will display these values. Simply click on the required option.

Finally if the user wishes to change the type of a defects, simply click on the relevant cell in the 'Defect Name' column. The dialog discussed in ***section 4.5*** will load. Unlock [via icon in top right corner] the defect list, select the desired defect and then change options as appropriate.

4.6.2 Block editing

Often a user will want to simultaneously set the values for a single parameter in a block of defects. For example, he may want to change the Y relative aperture from 0.5 to 0.7 for a number of TRA defects.

Simply select the block of cells in the appropriate column. A dialog will appear allowing the user to specify the new value for that parameter.

<table>
<tr>
<td>
</td>
<td>Defect type
This lists [left side] all defect types within the block of defects.
For the selected defect type in that list, you may choose to replace it with any one of the defects of the list on the right [these all have the same parameter types]
On OK the defect types are changed, but the actual parameters are left unchanged.</td>
</tr>
<tr>
<td>
</td>
<td>Wavelength #, Zoom # or Surf #
This dialog shows the limited number of options. Click on an option and the new value will be applied to the defects.</td>
</tr>
<tr>
<td>
</td>
<td>Ray Data, Target or Tolerance
[continuous range of variables]. The list shows the types of defect within the selected range. The user may choose to apply the change to all defects or to a selected type of defect.
The user may choose to set the parameter to a specific value, for all the updated defects.
Alternatively the user may choose to scale the existing parameter values by a specific amount.</td>
</tr>
</table>

4.6.3 Cut/copy/paste

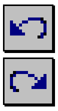

cut
copy
paste

The user may cut/copy and paste entire defects from/to the merit function.

To select a block of defects, move the cursor to the extreme left-hand column of the spreadsheet, over the number of the first defect in the block. Select the block of defects and release the mouse. The user may then cut or copy the block to the clipboard by pressing the appropriate button

To paste the contents, place the cursor in the row of the defect after/at which the defects on the clipboard are to be inserted. Click the paste button.

4.6.4 Undo/redo


undo
redo

Multi level undo/redo facilities exist in the merit function editor. Before each edit is implemented, the old merit function is recorded, and so is available for recall.

The stack of old merit functions is cleared when the user edits the lens, loads a new loads or clears the merit function.

4.7 Active/inactive defects

Not all defects listed in the editor are used in the merit function calculations. Obviously defects with zero tolerances [***section 4.1***] cannot be included.

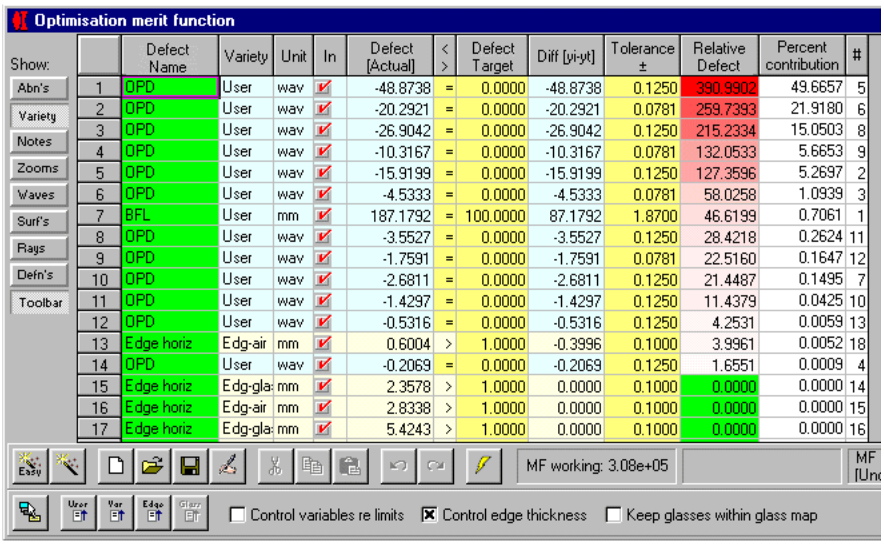
In WinLens it is also possible to define defects which have non-zero tolerances but are 'in-active'. Simply uncheck the check box option in the appropriate cell in the 'In' column in the spreadsheet. Reversing the process makes the defect active again. By default, all defects are active.

What is the point of this capability? Well, it is possible to use these 'in–active' defects as watchdogs to keep an eye on some aspect of system performance that the user is not happy about.

4.8 Sort defects by contribution

Normally the merit function spreadsheet shows the defects listed in numerical order. However it is very easy to sort the defects by size of contribution to the merit function.

Simply click on the header to the column 'Relative defect'. The defects will then be sorted with the largest contributions topmost. Note the extra column visible at extreme right, showing the normal sequence number of the defect.



	Defect Name	Variety	Unit	In	Defect [Actual]	< >	Defect Target	Diff [yi-yt]	Tolerance ±	Relative Defect	Percent contribution	#
1	OPD	User	wav		-48.8738	=	0.0000	-48.8738	0.1250	390.9902	49.6657	5
2	OPD	User	wav		-20.2921	=	0.0000	-20.2921	0.0781	259.7393	21.9180	6
3	OPD	User	wav		-26.9042	=	0.0000	-26.9042	0.1250	215.2334	15.0503	8
4	OPD	User	wav		-10.3167	=	0.0000	-10.3167	0.0781	132.0533	5.6653	9
5	OPD	User	wav		-15.9199	=	0.0000	-15.9199	0.1250	127.3596	5.2697	2
6	OPD	User	wav		-4.5333	=	0.0000	-4.5333	0.0781	58.0258	1.0939	3
7	BFL	User	mm		187.1792	=	100.0000	87.1792	1.8700	46.6199	0.7061	1
8	OPD	User	wav		-3.5527	=	0.0000	-3.5527	0.1250	28.4218	0.2624	11
9	OPD	User	wav		-1.7591	=	0.0000	-1.7591	0.0781	22.5160	0.1647	12
10	OPD	User	wav		-2.6811	=	0.0000	-2.6811	0.1250	21.4487	0.1495	7
11	OPD	User	wav		-1.4297	=	0.0000	-1.4297	0.1250	11.4379	0.0425	10
12	OPD	User	wav		-0.5316	=	0.0000	-0.5316	0.1250	4.2531	0.0059	13
13	Edge horiz	Edg-air	mm		0.6004	>	1.0000	-0.3996	0.1000	3.9961	0.0052	18
14	OPD	User	wav		-0.2069	=	0.0000	-0.2069	0.1250	1.6551	0.0009	4
15	Edge horiz	Edg-gla:	mm		2.3578	>	1.0000	0.0000	0.1000	0.0000	0.0000	14
16	Edge horiz	Edg-air	mm		2.8338	>	1.0000	0.0000	0.1000	0.0000	0.0000	15
17	Edge horiz	Edg-gla:	mm		5.4243	>	1.0000	0.0000	0.1000	0.0000	0.0000	16

To revert to normal order, simply click anywhere in another column.

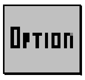
5 Optimisation options

The optimisation options include:

- *Optimisation strategy*
- *Separation limits*
- *Damping factor*
- *Glass 'contribution' function*
- *Update graphs/tables frequency*
- *Backup frequency*
- *Ending optimisation*

"Every program has at least one bug and can be shortened by at least one instruction - from which, by induction, one can deduce that every program can be reduced to one instruction which doesn't work."

- Anonymous



The results of an optimisation can be effected by a number of different options. Other options will determine when the optimisation ends and how often the graphs/tables are updated during optimisation.

All these options are available in a single multi-page dialog. This dialog is launched by clicking the 'Options' icon on the optimisation tab of the main toolbar.

This dialog has a number of pages covering different aspect of the optimisation process. In order, these pages are:

- *Optimisation strategy
- *Separation limits
- *Damping factor
- Glass 'cost' function
- Update graphs/tables frequency
- Backup frequency
- Ending optimisation

Any page may be viewed by clicking on the list at the right hand of the dialog. Any changes made are noted in the registry, and used until changed.

The *options are also saved in the lens file. By default they do not overwrite the options held in the registry. However, the user may choose to have automatic update of options on file load, by checking the check box on the right had side. This may be useful when repeating optimisation runs after a long time.

Finally it is also possible to reset the *options to defaults, by clicking the 'Defaults' button. These defaults are:

- Optimisation strategy: Simple
- Separations limits 1.0mm [with tolerance of 0.1]
- Damping factor Additive

5.1 Optimisation strategy

Optimisation involves some searching in variable space. This is done by varying the 'damping' factor and finding that which gives the smallest value for the merit function.

Each damping factor determines a slightly different direction [a different set of parameters] in the variable space. A line search method is also available which investigates each direction more thoroughly before choosing the final direction. This is more thorough but slower. Finally, the user may manually select the damping factor, if he so wishes.

The technical details of the line search strategy are discussed in ***section 12.3***.

5.2 Separations at limits

During optimisation it is important to control separations in each space, to make sure that the system is physically realisable. Both centre and edge thickness can be important.

On this tab the user may set default minimum centre & edge thickness values. The edge defaults will be used for all spaces[7], if the user has allowed for automatic control of edge thickness [***section 4.2.3.1***].

5.3 Damping factor

This is a technical option with two possibilities:

- Additive damping [default]
- Multiplicative damping

These are widely discussed in the literature, and are available for experimentation. It is said that multiplicative damping is particularly useful where aspherics are optimised.

5.4 Glass 'contribution' function

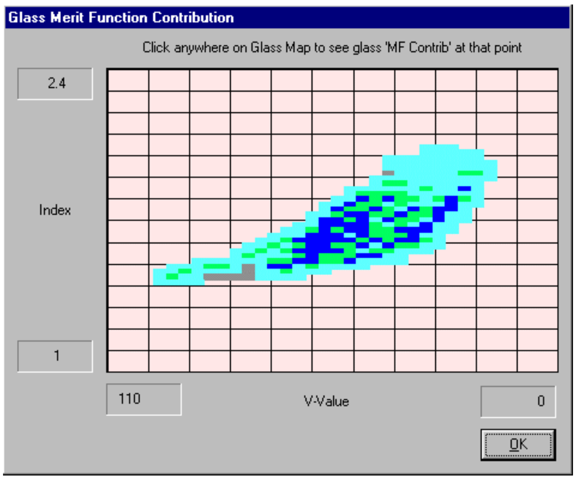
There are a limited number of real glasses; these all lie in one irregular block in n, V space. DLS optimisation must have continuous variables, so optimisation

[7] Except spaces where the user has deliberately defined edge separation limits in the variable editor [***chapter 3***]

'glasses' very rarely correspond exactly to a real glass. For the sake of matching, it is crucial that the optimised 'glasses' lie within the real glass region.

As a crude control the limits on n & V may be used. However WinLens Plus also offers automatic defects [***section 4.2.3.2***] for this purpose. For each variable glass a defect is added to the merit function. This defect assesses the value of the 'glass contribution function' for the n, V values of the glass.

A representation of the glass contribution function may be seen, by clicking the 'Show…' button on this page of the options dialog.

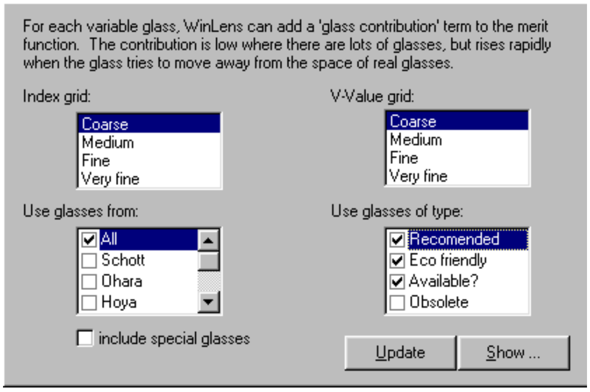


Clicking on the plot will show the contribution at that point in [n, V] space.

The function value is low (<.1) in regions where there are many glasses, is higher in regions of less glass numbers (<1) and rises rapidly in regions of no glasses, as the point moves further away from the real glass region.

The function is based upon an evaluation of glass numbers from glasses selected from the glass database in a grid of cells in n, V space. The function value at any point is obtained by bi-linear interpolation from the cell and its neighbours.

The user can determine [to some extent] the glass contribution function. He may specify coarseness of the grid, and also which glasses are to be used in the evaluation, by specifying manufacturer and glass category [recommended, eco-friendly etc]



When he is satisfied with his choices, he must press the 'Update' button, which will force a revaluation of the glass contribution function.

The 'Show…' button can then be clicked to display the modified function.

5.5 Update graphs/tables frequency

During the optimisation process, loaded graphs and tables may be updated at regular intervals. However this may cause the optimisation to really slow down. The user may choose when these solutions are saved:

- after optimisation ends
- after each cycle is complete
- after each 'good'[8] solution is generated
- after each new solution is generated

5.6 Backup frequency

During the optimisation process, solutions are saved at regular intervals. These solutions may then be reviewed afterwards [***chapter 7***]. The user may choose when these solutions are saved:

- after optimisation ends
- after each cycle is complete
- after each 'good' solution is generated
- after each new solution is generated

5.7 Ending optimisation

Optimisation will naturally end when minimum is reached. However the user may also elect to add the following termination conditions

- optimisation ends after prescribed number of cycles
- optimisation ends if merit function drops by less than x% over two cycles
- optimisation ends if merit function drops by less than y% within one cycles
- optimisation ends when all effective defects are within tolerance

The optimisation will automatically stop if it detects that it is at a minimum.

[8] a 'good' solution means that the merit function value for that solution is smaller than the merit function at the beginning of the cycle.

6 Control during the optimisation

- *Control options*
- *Feedback*

"In a few minutes a computer can make a mistake so great that it would have taken many men many months to equal it."

- Unknown

Once the optimisation has started, the solutions obtained depend only upon the optimisation algorithms, i.e. cannot be ‘steered’ by the user.

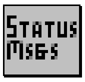
However, the user can pause, restart, & stop the optimisation at will. These functions are covered in this short chapter.

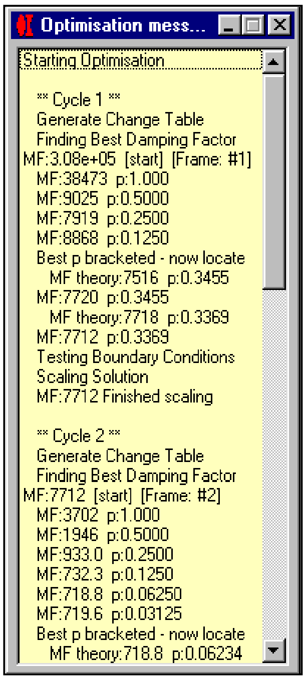
The buttons are held on the optimisation toolbar [***chapter 2***]. Although all are show enabled in the graphic above, in practise most will be disabled at any one time.

Function	**Event**
Go	Starts the optimisation, disables all other editors
Pause	Pauses the optimisation.
Continue	Resumes the optimisation
Stop	Stops the optimisation at the end of the current cycle. Re-enables editors
Stop!!	Stops the optimisation after the next solution. Re-enables editors

6.1 Feedback during optimisation

It is possible to view a ‘history’ of the optimisation, during and/or after the optimisation. This history is displayed in the status message form, as shown here.





To load the status message form, simply select the ‘Status messages’ option from the Optimisation menu, or click on the button [shown left] on the optimisation tab.

Looking at this provides a clear overview of what the optimisation is doing, and also how well it is progressing.

For example, if the merit function is hardly changing, then the user may wish to stop the optimisation and review the variables chosen or inspect the merit function.

Whenever a system is saved to disk [for video replay purposes] the message has [Frame #n] appended to the end.[9]

Once optimisation has finished, if you click on any message with the [Frame #n] then that solution will be reloaded for your inspection. This complements the ‘video’ replay system [***section 7.1***]

[9] The frequency of backup is set in the optimisation options dialog [***section 5.6***]

7 Replay after optimisation

- *Video buttons*
- *Bookmarking*

"The labelling of any engineering project with the epithet `simple' is more often than not the result of the feeling of confidence that springs from inadequate comprehension of the situation".

- SubG

In comparative trials of lens design software, it is always the skill of the designer which counts. The final polished solutions are the result of repeated trials and modifications, going back over old ground, branching off at some mid point and starting again. To a designer involved in this process, one of the most frustrating things in optimisation is seeing an interesting system flash by, to be lost forever.

In the light of this, simple tools have been developed to make easy the review/re-run process. These consist of:

- Optimisation video
- Bookmarking

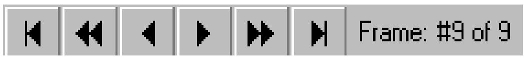
7.1 'Video'

During optimisation, the program automatically records some or all solutions to a dedicated folder, rather like a film or video.

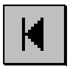
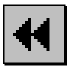
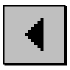
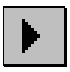
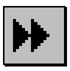
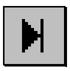
When the optimisation is completed, simple video controls are enabled. These allow the designer to re-run the optimisation in either direction, to jump to the start or end or to single step in either direction. Alternatively the designer can go straight to any recorded solution by clicking on the log spreadsheet[10].

Any of these solution, when recalled, can then be saved normally using the standard file options, bookmarked [see following section] subject to further analysis or even used as the starting point for further optimisation.

The user may determine when the solutions are saved to disk [***section 5.4***].



The buttons are held on the optimisation toolbar [***chapter 2***]. The toolbar also shows which frame or solution in the sequence is currently on display.

Function	**Event**
	Go back to start point [frame #1]
	Run through all solutions before the current solution until the start point is reached.
	Load previous solution in the optimisation sequence
	Load next solution in the optimisation sequence
	Run through all solutions after the current solution until the last solution in the optimisation is loaded.
	Load last system in optimisation

[10] If you look at the 'status message' form, you will see lines containing 'frame #n'. Each 'frame' is one of the solutions. If you click on one of these lines then that solution will be loaded.

7.2 Bookmarking

 bookmark system

 bookmark & notes

 review list of bookmarked systems

The 'video' temporarily notes solutions during one optimisation. However, once optimisation is restarted, the log is cleared, and the old intermediate solutions are overwritten.

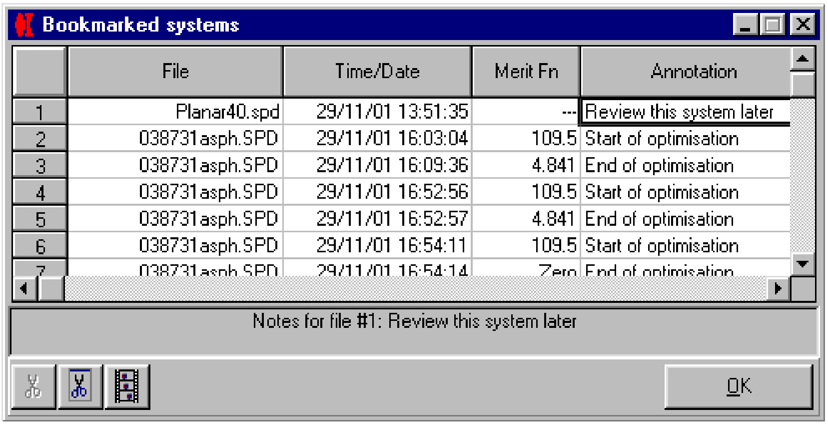
Therefore to preserve interesting solutions for longer-term inspection, a book-marking capability has been added to WinLens Plus.

When bookmarked, a copy of the current design is saved automatically to a dedicated folder. A list of all book-marked systems, showing time of creation, merit function value and any annotation is available. Then, also when required by the designer, any system in the list will be displayed.

During optimisation the first and last systems are automatically bookmarked. The designer can also manually bookmark a system at any time, not just during optimisation [using the buttons shown here at extreme left of the main toolbar].

To review bookmarked systems, click the third button shown. WinLens will then display the current list of bookmarked systems in a window. Each system has its own line.

To load a specific system simply click on that row. IN this way, you can easily look back, not just over the last optimisation, but over many past trials.



In this window you may also:

- Delete selected systems from the list
- Clear the entire list
- Cycle through the entire list, loading each in turn for a few moments.

Using the buttons at the bottom left of the window

8 How to

Notes on how to:

- *Optimise glasses*
- *Keep variables within limits*
- *Control edge thickness*
- *Optimise zoom systems*
- *Optimise aspherics*
- *Optimise systems with mirrors*

"Computer Science is no more about computers than astronomy is about telescopes."

- E.W Dijkstra

This chapter provides some brief notes and pointers to other sections of the manual. It deals with various tasks that some users may have to undertake.

8.1 Optimise glasses

Glasses are not simple variables, unlike radius or thickness. A glass has a refractive index which is a complex function of wavelength. In WinLens Plus, you can allow the index at the mid wavelength and/or the dispersion to vary. However at the end of optimisation, a real glass must be substituted.

The main problem is that there is only a limited range of real glasses available for use. We use 'normal' glass model for the variable material, but then have the problem of keeping that 'glass' within the real glass region.

You can allow the merit function to automatically add terms to keep variable glasses within the real glass region.

You can also place hard upper and lower limits on the index [if variable] and the dispersion [if variable]. These cannot be violated in optimisation.

Action	**Notes**
1 Define glass variables	***Section 3.1.3***
2 Apply glass limits	***Section 3.1.3*** [optional]
3 Request automatic glass control	***Section 4.2.3.2*** [optional] glass control function demo - section 5.4
You must control variables glasses in some way, so use either 2) or 3) or both	

Once optimisation is finished, you will want to find the nearest real glass. This can be achieved with the help of the 'Alternate glass finder' on the 'Database' tab. This will generate a sorted list of the real glasses and allow you to substitute one of these for the 'optimised' glass.

8.2 Keep variables within limits

To keep a variable within an upper and/or lower limit you must:

- either specify hard limit[s] in the variable editor
- or add defects to the merit function to make sure that the variables stay within the required range

The latter can be done manually by adding appropriate defects [***section 4.5***] from the 'design data' family [***chapter 11***]

8.3 *Control edge thickness*

It is very important to ensure that the results of an optimisation are physically viable! In particular edge thickness must be of the same sense as the axial separation in that space, i.e. the surfaces must not 'cross over' within the volume defined by the clear apertures on those surfaces!

You can of course manually add defects [***section 4.5***] from the 'edge control' family [***chapter 11***], for some or all spaces. However, WinLens Plus can do this for you automatically.

Action	Notes/manual section
1: Make sure default edge limits are correct	***Section 5.2*** [optimisation options] Look at the 'separations limits' page and make sure the values are ok
2: Request automatic edge thickness control	***Section 4.2.3.1*** [merit function editor] Check the 'Control edge thickness' box
If you follow this procedure, you will see that [toward the bottom of the merit function] there are new defects controlling minimum edge thickness	
3: inspect/alter tolerances on these new defects	***Section 4.3.1.2*** [merit function editor]

If you add/delete components or insert/delete surfaces in a component, the merit function will be adjusted automatically.

You may want to set an upper limit on the edge thickness in a space, but this is not usual, and hence must be done manually [***section 4.5***].

8.4 *Optimise zoom systems*

Special provision has been made for such systems. These provisions are discussed in sections listed below:

Provision	Notes
Zooming gap/defocus variables	***Section 3.1.2.2***. Allows user to set variables and limits on a zooming gap for some/all zooms. These variables/limits may vary independently between zooms
Setup merit function wizard	***Section 4.4.2.1***. Allows user to create merit function with same or different requirements for some or all zooms. [some zooms can be excluded]
User defined defects	***Section 4.6***. Most of the user-defined defects allow the user to specify the zoom to which they are attached. These can also be directly edited. Some defects allow control of the paraxial properties of groups of components [see ***page 69***]
Automatic edge thickness control	***Section 4.2.3.1***. When this option is selected, the merit function will add/maintain an edge thickness defect for each zoom for each zooming gap, allowing thickness control independently within each zoom.

If zooms are inserted/deleted, then WinLens Plus automatically maintains & updates the references within the defects to specific zooms.

8.5 Optimise aspherics

It may be that you know already which surface you wish to aspherise and what degree of aspheric is appropriate. In this happy case all you need to know is how to define those parameters as variable. This is covered in ***section 3.1.1.1***.

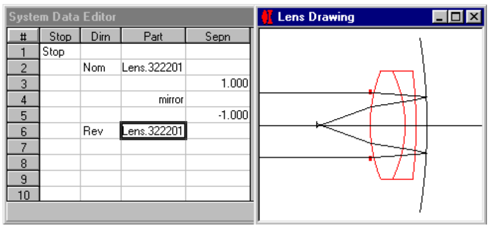
However, you may not know which surface to aspherise, which aspheric coefficients to vary or what aspheric definition is best. While WinLens Plus cannot give you a magic answer, WinLens Plus can help you come to conclusions.

Before you specify which parameters are variable, we suggest that you view the asphersation guide [***section 3.2.2***]. Press the update icon on that form and then review the results [this does a one-cycle optimisation of each parameter]. You can see how much the merit function drops for each surface. This provides a good starting point for your selection.

8.6 Optimise systems with mirrors

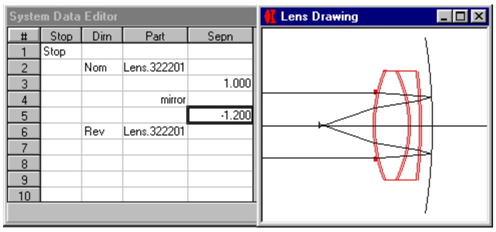
The main problem with optimising mirror systems is not the actual mirrors! It is components through which the light passes twice or more.

In terms of the design data the component will appear two or more times; although these are, in reality, one and the same component. A simple example is shown below:

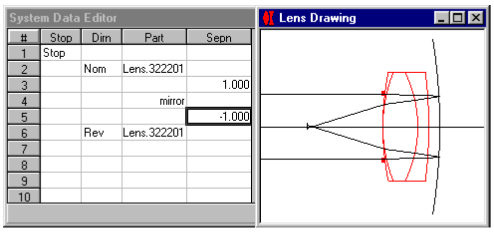


The 'two' components exactly overlay – this is crucial for a correct representation of reality. This is easy to achieve on setting up, but what about during optimisation?

The simple-minded user might allow gap 3 to vary or a curvature in the first component to alter, with the sad results shown below:



In neither case are the components identical. In the first the location is wrong, in the second the component prescription is different.

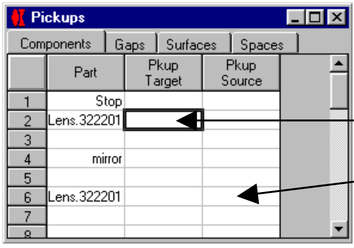


However this is easy to prevent using pickups. In this case we use two pickups:

- A component pickup
- A gap pickup

The component pickup ensures that all parameters within the two linked components are linked. The linkage depends upon the scale factor used in the pickup. In this case a value of –1 must be used [the negative value means that the component is reversed, and the value 1 enforces a unit scale factor].

The pickup editor can be accessed from the 'table' tab.

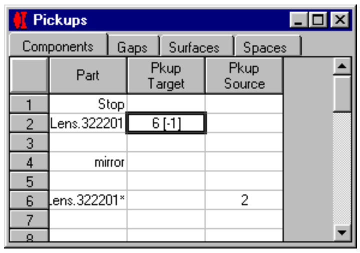


No pickups initially. To link the components with a pickup:

Mouse down in the 'highlighted cell'. [This is the 'source' component]

Drag and drop in the row for the target

Enter a value of –1 for the component scale factor



Completed component pickup.

Component 2 is the source of the pickup

Component 6 is the target of the pickup. The scale factor of –1, means all dimensions are the same, but the component is reversed.

A similar process is used for the gap pickup. Now if any parameters in the component/gap before the mirror are made, these will be duplicated in the component/gap after the mirror, thus keeping the proper synchronisation>

9 Examples

This chapter contains:

- *Example 1 – location of essential optimisation tools*
- *Example 2 – defining variables & creating a default merit function*

"I do not fear computers. I fear the lack of them."

- Isaac Asimov

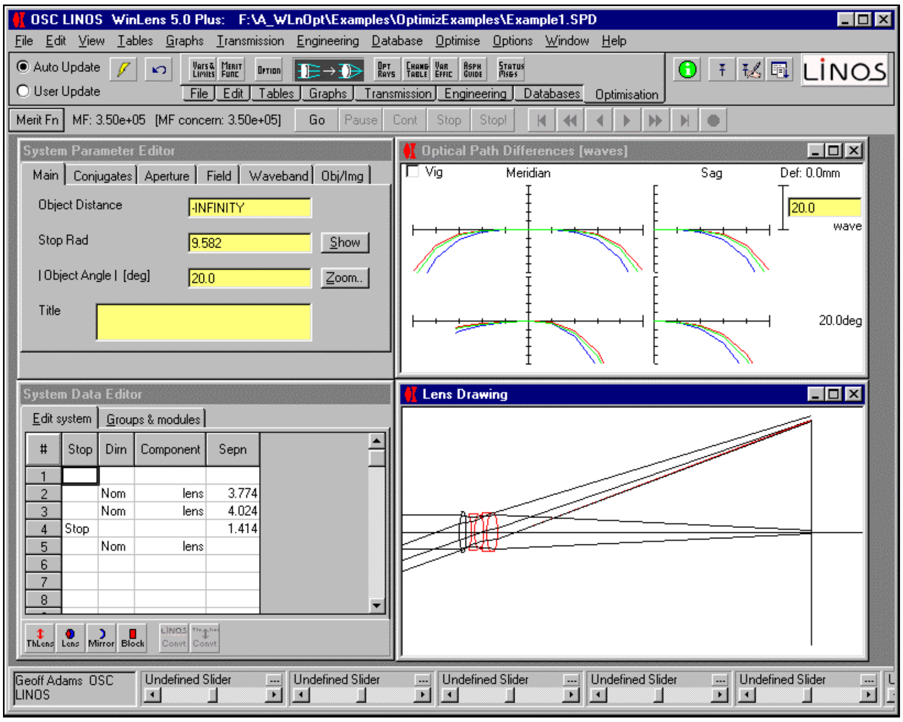
This chapter is contains two examples – the first is just to familiarise you with the location of the controls. The second is more complex and includes setting up a default merit function and selecting some variables.

We assume that you probably have not read the rest of the manual, so there will be lots of graphics! Most of these are representative only; you may have moved editors/tables & graphs around, or have different tables & graphs loaded to suit your own purposes.

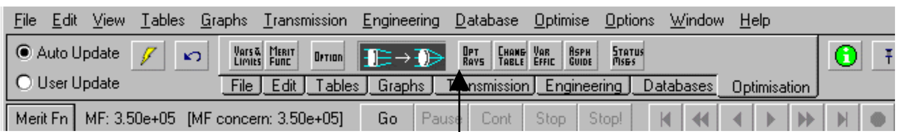
9.1 Example 1: basic tour

Load the file 'Example 1.SPD'. Below we show a full screen display with:

- Optimisation tab selected in main toolbar at centre top of screen.
- Lens drawing clearly showing badly aberrated rays
- OPD plot [this is just one of the many graphs available]. These plot[s] will be updated during optimisation, to help you monitor progress[11]



In order to start the process, the optimisation control bar must be made visible:



Control bar, with:

- Current merit function
- Start [stop…] buttons
- Replay controls

Press this icon to make control bar visible

load options dialog

Don't yet press the 'Go' button to start the optimisation. We want you to set a couple of the optimisation options first. You won't need to do this each time!

[11] Of course, more graphs/tables loaded means that extra calculations will have to be performed, which slows down the optimisation. The user may therefore limit the number of graphs/tables and/or reduce update frequency [***section 5.4***]

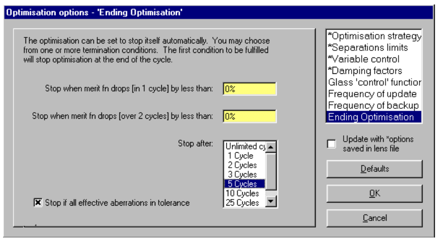
So click on the ‘Options’ button on the ‘optimisation tab [or select the ‘Options’ from the Optimise menu].

Click the ‘Defaults’ button first. This is just to make sure that the main options are the same as when we ran the optimisation.

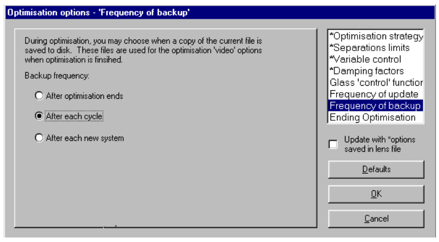
Next, the optimisation options has several pages [see ***section 5***], which can be selected from the list at the right hand side. We are also going to:

- specify conditions for ending the optimisation
- specify frequency of backup during optimisation

Go to the tabs labelled ‘Ending optimisation’ and ‘Backup frequency’ respectively and set them as shown below.



Optimisation to end after 5 cycles



Backup made after every cycle [these will be used in the ‘video’ afterwards]

[If you want, you could also make sure that the ‘Update frequency is ‘after every cycle’ or ‘After every new system’. This determines when the graphs/tables are updated]

When these are set, click the ‘OK’ button to accept the changes and return to the main program.

start optimisation

Now start the optimisation and watch the lens drawing and TRA plot [or any other graphs] as the optimisation proceeds. Initially you will see large changes in performance curves, but as the optimisation gets nearer to a minimum so the changes get smaller. When optimisation stops, you should see something like:

Merit function falls

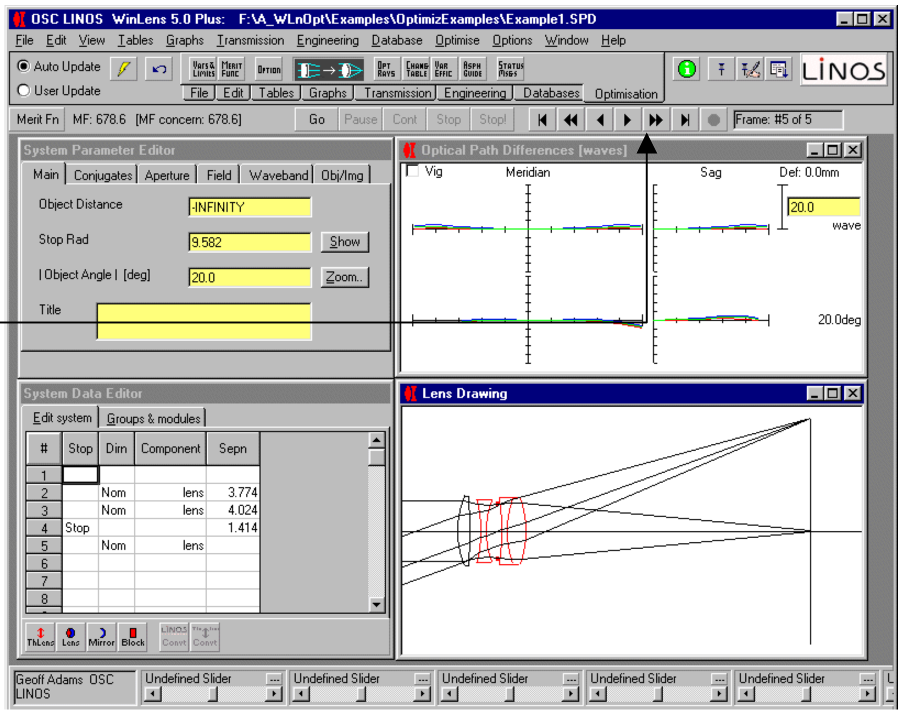
- from 3.5×10^5
- to 678.6

during 5 cycles

Video controls enabled when optimisation ends.

A sequence of solutions is saved in optimisation.

Use these controls to review those solutions.



Note how the performance has improved radically after optimisation; the OPD dropping from over 20 waves of defect to under 4 waves!

That ends the first example.

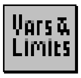
9.2 Example 2: simple merit function and variables

In order to optimise an optical system, you need to have some merit function and you need to have some variables. In this example you will learn how to:

- set variables [***chapter 3***].
- how to create a merit function [***chapter 4***] using a wizard [***section 4.4***].

You will also see how to review status messages during optimisation.

Start by loading 'Example 2.SPD'[12] .

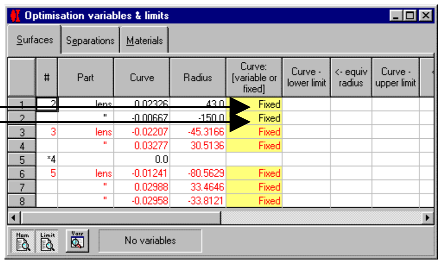


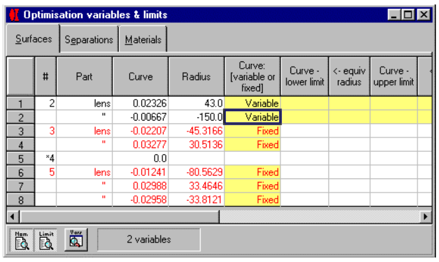
load variable editor

Display the variable editor. This is a tabbed form; each tab being for different classes of parameter. We want to make some radii variable, so select the tab labelled 'Surfaces':

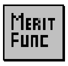
Before: all radii are fixed.

Click in cells to make parameters variable



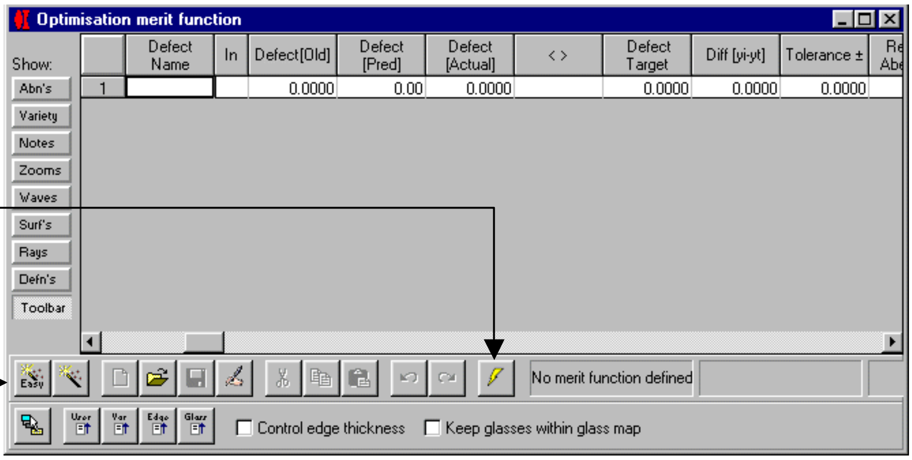


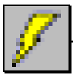
After: radii on surfaces 1 & 2 are variable.



load merit function editor

Now that we have two variable radii, we need a merit function. To create or review or edit the merit function, you must load the merit function editor. This is a sophisticated tool with many capabilities, but in this example we will do the least possible!





Click to update merit function [if any!]

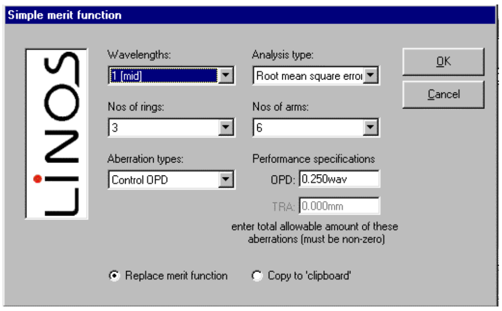
Click to start quick setup wizard

A detailed discussion of all the editor function, including direct editing, cut, copy & paste, undo/redo etc, etc can be found in ***section 4.3*** & ff.

To create a simple merit function, click the quick setup wizard icon. This will load the quick setup wizard.

[there is another full featured wizard available, but we will not be using that for this example]

[12] This is the same lens as held in Example1.SPD, but without a merit function and with no variables.



The ray ring type pattern is described in detail in ***section 4.4.2.3***

This will create a simple merit function which will aim for a diffraction limited system and controlled back focal length

Click the 'OK' button

The merit function editor will now show the simple merit function created for this system. The set of defects is based upon the ray ring pattern described in ***section 4.4.2.3*** above.

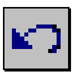
Target BFL value to be changed manually to a value of 100mm. Simply click on this cell and enter 100

	Defect Name	In	Defect [Actual]	< >	Defect Target	Diff [yi-yt]	Tolerance ±	Relative Defect	Percent contribution
1	BFL	✓	194.29272	>	194.2927	0.00000	1.94000	0.0000	0.0000
2	OPD	✓	-16.94419	=	0.0000	-16.94419	0.12500	135.5535	5.2847
3	OPD	✓	-4.84626	=	0.0000	-4.84626	0.07813	62.0322	1.1067
4	OPD	✓	-0.22220	=	0.0000	-0.22220	0.12500	1.7776	0.0009
5	OPD	✓	-51.69249	=	0.0000	-51.69249	0.12500	413.5399	49.1848
6	OPD	✓	-21.65278	=	0.0000	-21.65278	0.07813	277.1556	22.0924
7	OPD	✓	-2.93950	=	0.0000	-2.93950	0.12500	23.5160	0.1590
8	OPD	✓	-28.97986	=	0.0000	-28.97986	0.12500	231.8389	15.4586
9	OPD	✓	-11.22388	=	0.0000	-11.22388	0.07813	143.6657	5.9361
10	OPD	✓	-1.58002	=	0.0000	-1.58002	0.12500	12.6402	0.0460
11	OPD	✓	-4.91020	=	0.0000	-4.91020	0.12500	39.2816	0.4438
12	OPD	✓	-2.43282	=	0.0000	-2.43282	0.07813	31.1401	0.2789
13	OPD	✓	-0.66428	=	0.0000	-0.66428	0.12500	5.3143	0.0081

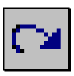
Before you do anything, just look at the target values, and especially non zero values! It is good practise to do this always, just to make sure that the are no unexpected values.

In this case the BFL has a target value of 194.293mm. This is taken directly from the system value when the wizard was launched. However, we want a value of 100mm for the BFL. So position the mouse cursor in the cell and enter the value 100.00. Note how the defect contribution and therefore the merit function value increases.

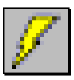
Now you have both variables and a merit function. Let's use some of the editing capabilities – try out the undo-redo buttons.


undo

Click once – this will undo the change to the BFL target
Click once again – this will undo the action of the wizard
[Note the tooltip shows the nature of the undo about to be made]


redo

Click once – this will redo the merit function change
Click once again to redo the change to the BFL target


update

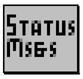
At this point, if you look at the 'Defect Name' column you will see that all the defect names have a grey back colour. This means that they are unverified. You will also note that the contributions column is hidden.

Click on the 'update' button to:

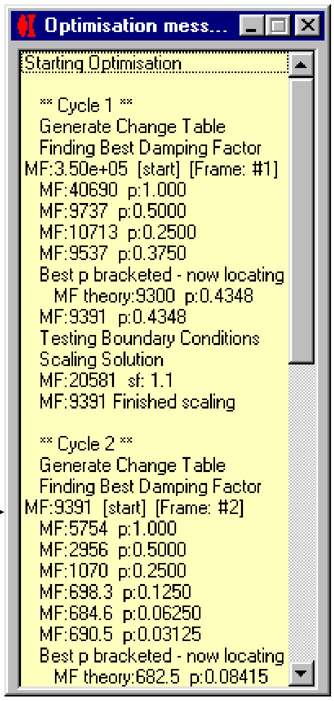
- Verify defects are valid
- Calculate & display defect values & contributions
- Calculate & display merit function value

As you can see, the undo-redo keeps track of multiple changes to the merit function – enabling you to rectify mistakes easily!

Finally, before we start the optimisation, lets load the status message window. This is NOT needed to do an optimisation – it will work exactly the same. However, for those of you who are interested, this gives a clearer picture of what is going on in the optimisation.



Icon to show/hide status messages form is on 'optimisation' tab.



These solutions have been saved to disk.

Click on one of these 'Frame' messages to re-load that solution

This form can be loaded at any time to review last several hundred status messages [even after optimisation is finished]

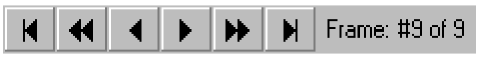
This form shows the process of optimisation for each cycle until optimisation ends. [here we see messages from the first two optimisation cycles]

Whenever a system is saved to disk [for video replay purposes] the message has [Frame #n] appended to the end.

Once optimisation has finished, if you click on any message with the [Frame #n] then that solution will be reloaded for your inspection. This complements the optimisation 'video' replay system [***section 7.1***]

Now start the optimisation, exactly as described in example 1, and watch the optimisation progress.

Afterwards try using the 'video' replay buttons to move back and forward through the optimisation solutions. Alternatively use the status message form as described above



10 Glossary

- *Simple definitions of frequently occurring terms*
- *Extended definitions of some of the above*

"To err is human, but to really foul things up requires a computer."

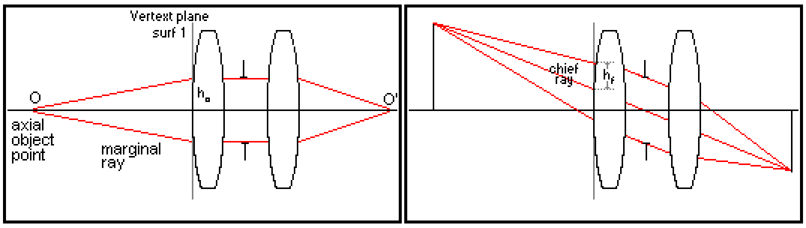
- Farmers' Almanac

This table contains brief definitions of terms used in optimisation. Some of these are illustrated overleaf.

Term	Definition
Defect	Any measure of the system performance, such as TRA or OPD – this can include many quantities not strictly thought of as defects, such as focal length.
Defect contribution [Relative defect]	Defined as: (defect - target) / tolerance [a.k.a. relative defect]
Defect space	Multi-dimensional space, with one axis for each of the defect terms used to make up the merit function
Defect target	Desired value for a defect, often but not always zero. This target may, of course, not actually be achievable!
Defect tolerance	Acceptable difference between the actual defect value and the defect target value. This [when non-zero] is used to weight the defect contribution within the merit function.
Active defect	A defect which is ticked in the 'In' column of the merit function editor. Only active defects [with non zero tolerances] are used in the working merit function, i.e. only these defects are used to control the optimisation direction.
Change Table	A spreadsheet where: - Each row contains data for a single defect, j - Each column contains data for a single variable, k The data being difference in defect j, due to a small change in variable k. The change table is calculated and used in each optimisation cycle. A change table may also be generated for a range of parameters [e.g. all radii] to help in the selection of variables from that list of parameters
Defect vector	A vector in defect space linking the nominal system defects with the desired [or target] defect values.
Merit function	The merit function is used to generate a single number which summarises the performance of the system. The merit function is composed of a set of defects with particular targets & tolerances. The value of the merit function is defined as: (Sum (defect contributions)^2)
Relative aperture coord	A way of defining a ray in a pattern associated with a chief ray from a particular object point. Provides a normalised measure of the ray co-ordinate on the vertex plane to the first surface relative to the chief ray from the same object. See next page
Relative field coord	A way of defining a particular object, where the axial object has a REC of 0 and the full field object height has a REC of 1.
Ray Pattern	A set of rays associated with a chief ray. These are used to probe the performance of the system for a particular object point. Many patterns have been used in the past. In WinLens we offer two distinct patterns. See next page.
Variable	One of the system design parameters [such as radius, separation, index etc], which is allowed to be changed by the optimisation algorithm
Variable change vector	A vector in defect space [for one variable], linking the nominal defect values with the new values calculated after making a small change to that parameter alone.
Variable efficiency [VE] [***Section 12.2***]	A measure of the likely value of a parameter in reducing the merit function. It is defined as: 100 x Cosine(θ) where θ is the angle between the defect vector and the variable change vector. If these vectors are perpendicular [VE=0], then that parameter is very unlikely to reduce the merit function and hence is inefficient. By contrast, if the vectors are parallel or anti-parallel then changing that parameter will move the solution toward the desired target

Extended definitions of some glossary terms:

Relative aperture co-ordinates:
The relative aperture co-ordinate is set from the paraxial ray values at the vertex plane to the first surface.



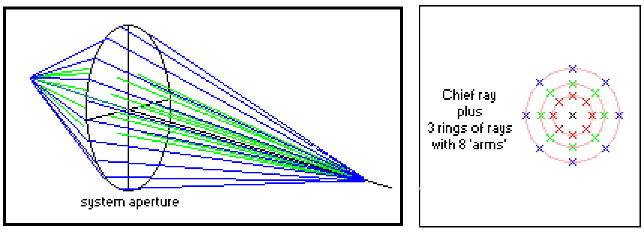
It is defined as h_f/h_o, where the these terms are shown in the diagram above.

Ray patterns:
In order to fairly assess the performance of a lens it is essential to trace many rays from several object points [on and off axis].

Of course the user may set up whatever defects [& hence rays] that he wishes. Typically however he will tend to use some standard pattern of rays. This pattern will tend to be duplicated for each object point of interest.

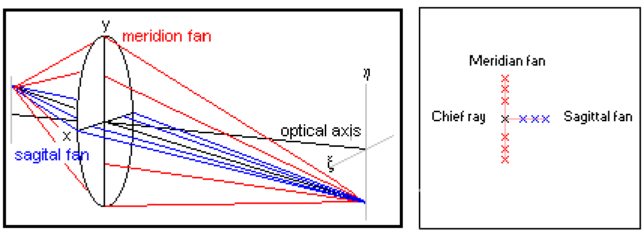
The setup wizard will generate sets of defects, which are based upon standard sets of rays. These particular patterns are widely used in the industry and have proved their worth over many years.

1) Ray Rings.
This involves tracing rings of rays concentric to the chief ray. Each ring contains the same number of rays. The spacing of these rings is determined by complex theory. The user chooses number of rings and number of 'arms' in the pattern



The theory behind this is discussed in detail in:
GW Forbes 'Optical system assessment for design: numerical ray tracing in the Gaussian pupil, JOSA p1943, Nov 1988

2) Ray Fans [Imperial college]
The ray fan pattern is more empirically based. It involves tracing fans of rays in the meridian and sagittal planes about each chief ray.



11 Defect definitions

Definitions of defects available to be used in the WinLens Merit function, arranged by family:

- *Conjugate parameters*
- *Field parameters*
- *Aperture parameters*
- *Paraxial parameters*
- *Paraxial ray data*
- *Seidel aberrations*
- *Real ray data*
- *Design data*
- *Edge thickness*
- *Combination*
- *Groups/Modules*

"Optical designers always want what you have not written, and are never grateful for what you have written."

- software engineer

Defects arranged by family:

Family	Defect	Definitions & Notes
Conjugate parameters		Uses values displayed in system parameter editor
	Obj Dist	Distance from object to first surface
	Img Dist	Distance from image to last surface
	L	Distance from object to principle plane P
	L'	Distance from image to principle plane P'
	Mag	Magnification
	Track	Distance from object to image
Field parameters		Uses values displayed in system parameter editor
	Obj angle	Object angle [arctan(object height/L)]
	Obj height	Object height
	Img angle	Image angle [arctan(image height/L')]
	Img height	Image height
Aperture parameters		Uses values displayed in system parameter editor
	Obj NA	Object space numerical aperture
	Img NA	Image space numerical aperture
	Stop rad	Radius of stop in system
	Entrance beam rad	Radius of axial beam at first surface
	Exit beam rad	Radius of axial beam at last surface
	F.Nos	Image Space Index/(2*Image NA) [when object at infinity]
	Entrance pupil rad	Radius of entrance pupil
	Exit pupil rad	Radius of exit pupil
	F.Nos effective	Image Space Index/(2*Image NA)
Paraxial parameters		Various useful paraxial quantities
	EFL	Effective focal length [f']
	BFL	Back focal length - distance from last to image surface
	Length	Distance from first to image surface
	Width	Distance from first to last surface
	Entr Pupil	Position of entrance pupil wrt first surface
	Exit Pupil	Position of exit pupil wrt last surface
	Front Focus	Distance of front focus from first surface
	Rear Focus	Distance of rear focus from last surface
	Pr. Plane 1	Distance of object side principle plane from first surface
	Pr. Plane 2	Distance of image side principle plane from last surface
	Nodal point 1	Distance of object side nodal point from first surface
	Nodal point 2	Distance of image side nodal point from last surface
Paraxial rays data		Results from paraxial ray trace
	H - axial	Ray height of axial marginal ray
	U - axial	Ray 'angle' of axial marginal ray [after at i^{th} surface]
	H - chief	Ray height of chief ray
	U - chief	Ray 'angle' of chief ray [after at i^{th} surface]

Family	Defect	Definitions & Notes
Seidel aberrations		
		First set of defects are the raw seidel coefficients [as noted], summed over entire system
	SphAbn	Sum of Spherical aberration coefficients [SI]
	Coma	Sum of Coma coefficients [SII]
	Astigmatism	Sum of Astigmatism coefficients [SIII]
	PtzCv	Sum of Petszval curvature coefficients [SIV]
	FldCv	Field curvature [(2SIII + SIV)/2]
	Distortion	Sum of Distortion coefficient [SV]
	LongChrAbn	Sum of Longitudinal Chromatic Aberrations coefficient [CI]
	LatChrAbn	Sum of Lateral Chromatic Aberrations coefficient [CII]
	S’ndry Spec	Sum of Secondary spectrum coefficients
	Pupil SphAbn	Sum of Pupil spherical aberration coefficients [SI’]
	Pupil Coma	Sum of Pupil coma coefficients [SII’]
	Pupil Distortion	Sum of Pupil distortion coefficients [SV’]
		Next set of defects are the raw seidel coefficients, summed from surfaces n to m
	SphAbn [surf]	Spherical aberration [SI]
	Coma [surf]	Coma [SII]
	Astigmatism [surf]	Astigmatism [SIII]
	PtzCv [surf]	Petszval curvature [SIV]
	Distn [surf]	Distortion [SV]
	LongChrAbn [surf]	Longitudinal Chromatic Aberrations [CI]
	LatChrAbn [surf]	Lateral Chromatic Aberrations [CII]
	S’ndry Spec [surf]	Secondary spectrum
	Pupil SphAbn [surf]	Pupil spherical aberration coefficients [SI’]
	Pupil Coma [surf]	Pupil coma coefficients [SII’]
	Pupil Distortion [surf]	Pupil distortion coefficients [SV’]

Note: all monochromatic aberrations may be evaluated at **any** one of the standard wavelengths [not necessarily the main (mid) wavelength].

Family	Defect	Definitions & Notes
Real ray data		Data extracted from rays required by merit function. Most are for data in image space. Those marked 'Ray: ' return ray values at/after specified surface.
	(T+S)/2	Average distance of astigmatic surfaces from image surface
	T-S	Horizontal distance between T&S surfaces along chief ray
	T	Horizontal distance T –> image surfaces along chief ray
	S	Horizontal distance S –> image surfaces along chief ray
	Distort	Distortion - chief ray
	Distort f-theta	F Theta distortion - chief ray
	Distort ang	Angular distortion - chief ray
	Distort ang f-theta	Angular f Theta distortion - chief ray
	TRA-X	X component of TRA [all TRA include any defocus effects]
	TRA-Y	Y component of TRA
	TRA-R	Magnitude of TRA [sqr(TRAX^2+TRAY^2)]
	TRA-X ang	X component of angular TRA
	TRA-Y ang	Y component of angular TRA
	TRA-R ang	Magnitude of angular TRA
	OPD	Optical Path Difference
	CPD	Conrady Path Difference [between λ_{long} & λ_{short}]
	CPD2	Conrady Path Difference [between λ_{mid} & λ_{short}]
	Sine condition	Sine condition for ray from axial bundle
	Isoplanatism	Isoplanatism for ray from axial bundle
	Ray: X [image]	X co-ordinate of ray at image surface
	Ray: Y [image]	Y co-ordinate of ray at image surface
	Ray: Z [image]	Z co-ordinate of ray at image surface
	Ray: N [Object]	N direction cosine in object space [useful for telecentricity]
	Ray: N' [image]	N direction cosine in Image space [useful for telecentricity]
	Ray: X [surf]	X co-ordinate of ray at i^{th} surface
	Ray: Y [surf]	Y co-ordinate of ray at i^{th} surface
	Ray: Z [surf]	Z co-ordinate of ray at i^{th} surface
	Ray: R [surf]	Ray 'height' at i^{th} surface [sqr(RayX^2+RayY^2)]
	Ray: L' [surf]	L direction cosine of ray after i^{th} surface
	Ray: M' [surf]	M direction cosine of ray after i^{th} surface
	Ray: N' [surf]	N direction cosine of ray after i^{th} surface
	Ray: segment	Length of ray segment from i^{th} surface to next surface
	Ray: CosI	Cosine of angle of incidence of ray before i^{th} surface
	Ray: CosI'	Cosine of angle of incidence of ray after i^{th} surface

Notes:
CPD, CPD2: Conrady aberrations - method of calculating the chromatic aberration along a ray. The ray only needs to be traced at one wavelength, using data from the standard OPD calculations and the glass dispersions. See Welford 'Aberrations of optical systems'. Adam Hilger, ISBN 0-85274-564-8

Family	Defect	Definitions & Notes
Design data		
	Radius	Radius of curvature of i^{th} surface
	Curvature	Curvature of i^{th} surface
	RadiusX	Radius of curvature i^{th} surface in x direction [toroid surface]
	Sepn	Axial separation after i^{th} surface
	RefIndex	Refractive index at j^{th} wave after i^{th} surface in k^{th} zoom
	V-Value	V-value $[(n_{mid}-1)/(n_{short}-n_{long})$
	Dispersion	$n_{short}-n_{long}$
	Rel PartialDisp	Relative partial dispersion $(n_{short}-n_{long})/\ (n_{short}-n_{mid})$
	Index max	Refractive index after i^{th} surface: max for all waves/zooms
	Index min	Refractive index after i^{th} surface: min for all waves/zooms
	Clear Radius	Radius of aperture limiting rays on i^{th} surface
	Sag	Sag of i^{th} surface at clear radius
	Conic Const	Conic constant
	A4 or B4 or C4	Aspheric coefficient of i^{th} surface [standard, normalised or Zernike]
	A6 or B6 or C6	"
	A8 or B8 or C8	"
	A10 or B10 or C10	"
	A12 or B12 or C12	"
	Bend	Bending of space between surfaces I to I+1. Defined as $(Cv_1 + Cv_n) / (Cv_1 - Cv_n)$
	Power	Power [1/efl] of surfaces n to m
	Defocus	Defocus for i^{th} zoom
Edge thickness		
	Edge full [zoom]	Total distance from surf i to surf i+1 between the heights of clear radius (surf i) & clear radius (surf i+1) respectively for a specified zoom [useful for zoom gaps]
	Edge full	Total distance from surf i to surf i+1 between the heights of clear radius (surf i) & clear radius (surf i+1) respectively
	Edge Horiz [zoom]	Horizontal part of Edge full [zoom] defect. **Used as automatic edge control defect**
	Edge Horiz	Horizontal part of Edge full defect. **Used as automatic edge control defect**
	Edge horiz n->n+1	Horizontal distance from surf i to surf i+1 at the height of clear radius (surf i)
	Edge horiz n->n+1	Horizontal distance from surf i to surf i+1 at the height of clear radius (surf i+1)
	Edge horiz max [zoom]	Max value of edge horiz in all spaces for specified zoom
	Edge horiz min [zoom]	Min value of edge horiz in all spaces for specified zoom
	Edge horiz max	Max value of edge horiz in all spaces for all zooms
	Edge horiz min	Min value of edge horiz in all spaces for all zooms

Note: the Clear Radius defect forces tracing of full ray fans at each solution and can therefore slow optimisation down considerably!

Family	Defect	Definitions & Notes
Combination		All these defects must refer to earlier defects in the merit function
	Add	p.abn_j + y.abn_k [where p & q are user defined values and abn_j & abn_k are i^{th} & j^{th} defects]
	Subtract	p.abn_j - y.abn_k
	Multiply	p.abn_j x y.abn_k
	Divide	p.abn_j / abn_k
	Sum	Sum(abn_i) for i = j to k
	Sum Squares	Sum(abn_i)^2 for i = j to k
	RMS	RootMeanSquare of abn_i for i = j to k
	RMS TRA	RMS of any TRA-X/TRA-Y defect's between abn_j & abn_k
	RMS TRA X	RMS of any TRA-X defect's " "
	RMS TRA Y	RMS of any TRA-Y defect's " "
	RMS TRA ang	RMS of any TRA-X ang/TRA-Y ang defect's " "
	RMS TRA X ang	RMS of any TRA-X ang defect's " "
	RMS TRA Y ang	RMS of any TRA-Y ang defect's " "
	RMS OPD	RMS of any OPD defect's " "
	Mean	Mean of abn_i for i = j to k
	Mean [abs]	Mean of abs(abn_i) for i = j to k
	Max	Maximum value of any abn_i in range i = j to k
	Max [abs]	Maximum value of any abs(abn_i) in range i = j to k
	Min	Minimum value of any abn_i in range i = j to k
	Min [abs]	Minimum value of any abs(abn_i) in range i = j to k
	Abn Type Sum	Sum of all abn_i of specified type in range i = j to k
	Abn Type RMS	RMS of all abn_i " "
	Abn Type Mean	Mean of all abn_i " "
	Abn Type Mean [abs]	RMS of all Abs(abn_i) " "
	Abn Type Max	Max value any abn_i " "
	Abn Type Max [abs]	Max value any Abs(abn_i) " "
	Abn Type Min	Min value any abn_i " "
	Abn Type Min [abs]	Min value any Abs(abn_i) " "
	Abs	Absolute value of j^{th} defect
	Square Root	Square root of j^{th} defect
	Square	Square of j^{th} defect
	Power	(j^{th} defect)^p
	Sin	Sine of j^{th} defect. [parameter assumed to be in radians]
	Cos	Cosine of j^{th} defect
	Tan	Tangent of j^{th} defect
	ArcSin	Arc sine of j^{th} defect – return value in radians
	ArcCos	Arc cosine of j^{th} defect
	ArcTan	Arc tangent of j^{th} defect

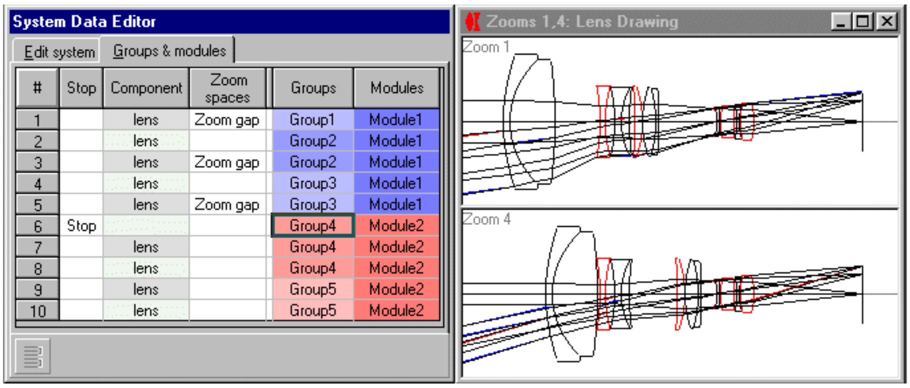
Family	Defect	Definitions & Notes
Groups/Modules		These defects refer to groups & modules as defined in the system data editor. **See notes below**
	Group EFL	Efl of the specified group [for specified zoom/wave#]
	Group BLF	Bfl of the specified group
	Group N[i]' – N[i+1]	Distance between principle plane P' of i^{th} group, and principle point P of following group
	Module EFL	Efl of the specified module [for specified zoom/wave#]
	Module BLF	Bfl of the specified module
	Module N[i]' – N[i+1]	Distance between principle plane P' of i^{th} module, and principle point P of following module

Rule1: A group is defined as one or more adjacent components. Groups cannot cross zoom spaces.
Rule2: A module is defined as one or more adjacent groups. Modules can cross zoom spaces.

Groups & modules are shown in the system data editor. In WinLens Plus, this has been modified with a tabbed interface. The first tab shows the old familiar components & spacings. The second tab shows the components and the groups/modules.

When a system is loaded WinLens automatically assesses the components and spaces and assigns each to a group. So for a single zoom system, all components are in one group, and there is one module. For a zoom system with one or more zoom gaps, there will be several groups, but still in one module.

To define a group: select one or more adjacent components in the 'component' column and click the 'grouping' icon at lower left.
To define a module: select one or more adjacent groups in the 'group' column and click the 'grouping' icon again.



Here we show a zoom system, with the default groups slightly modified.

Groups & modules do not impact on raytracing in any way. However in the paraxial component table and the Seidels table, the user may elect to see the data arranged surface by surface [default], by component, by group or by module.

Some defects, listed above, specific to groups/modules, have been created. These defects are useful in the optimisation of zoom systems, when trying to maintain focus through zoom. The back focal length of the system depends upon the power [efl] of each group and the separation of nodal points in adjacent groups. If the starting system has the desired behaviour, then by controlling these properties to the starting values, then the system will maintain zoom tracking during optimisation.

[An alternative procedure is to just put a heavy weight on zoom tracking during optimisation. This has the advantage that it allows individual groups to change so the software can discover new zoom constructions]

12 Appendix: Optimisation theory

This appendix covers optimisation theory as used in WinLens. Many papers and books have been written in the last four decades on this subject, so we just provide a very brief introduction.

The designer must have some sort of starting design, either derived from first principles, taken from the literature or an existing design.

An optical system is defined by many parameters, such as radius, separation, index etc. Small changes to the values of these can make significant differences to the performance of the system.

The designer must allow some of these parameters to vary during optimisation in order to improve the system. For these n variables, let us define:

- **initial values** of these variables as: $X_1, X_2, \ldots X_n$.
- **changes** in the variables as: $x_1, x_2, \ldots x_n$.

The performance of the system can be assessed by a whole set of measures, ranging from the paraxial quantities, such as back focal length, through Seidel aberration to real ray based quantities such as Isoplanatism, TRA, OPD etc. Let us define these m measures as:

- defects: $y_1, y_2, \ldots y_m$
- in particular, defects of the starting system: $y_{01}, y_{02}, \ldots y_{0m}$

Typically, there will be a great many more defects than variables, i.e. m>n.

The designer will have target values, $y^t_1, y^t_2, \ldots y^t_m$, for these defects, which are derived from the performance requirements. Frequently, but not always, these target values will be zero.

Almost invariably, the defects of the starting system will be much larger than the target values. The task of the designer is therefore to change the variables[13] in order to reduce the actual defects to the target values.

Usually, however, there are no simple links between one variable and one or more defects. To achieve the desired balance, multiple changes have to be made to the variables. This is beyond the designer, except in the simplest systems, and therefore requires some automatic optimisation process.

For the purposes of optimisation, we need to combine all these disparate defects into a single function – the merit function. For each defect we define a contribution, f_i, where:

$$f_i = w_I \times (y_i - y^t_i) \qquad \text{[eqn 1.1]}$$

where w_i is a weight[14] on that defect – a means of expressing the relative importance of the different defects. A contribution from the nominal system is denoted by f_{0i}. These form a vector that is often known as the defect vector.

The merit function is then defined as:

$$\varphi = \sum_{i=1}^{m} f_i^2 \qquad \text{[eqn 1.2]}$$

The smaller the merit function, the closer it is to the target values.

In the following sub sections we will discuss:

- Damped least squares optimisation
- Normalisation
- Line search strategy

[13] Part of the skill of the designer lies in the selection of variables. This is not covered in the appendix, but we provide tools to help with this task [***section 3.2***]

[14] In WinLens w_i = 1 / tol_i where tol_i is the acceptable range of variation of an aberration about the target value. For example, we may require a system to have a focal length of 100m ± 2%. In this case the tolerance, tol_i, is ± 2mm.

Damped least squares is the standard method of optimisation. In this section we shall show how it is derived [and see why that particular name is used!]

12.1 Standard optimisation

If contours of the merit function are plotted in n dimensional variable space then the result is almost invariably a highly complex folded structure with long narrow valleys. The starting point is usually somewhere high up and far away from these valleys. Our aim is to get down into a valley and reach its bottom.

First we need to lay out some maths we will use later.

We know that the defects change with the variables. Since these are, in general, continuous functions, we can expand each in a Taylor series about the starting values, i.e.:

$$f_i = f_{oi} + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} x_j + \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial f_i}{\partial x_j} \frac{\partial f_i}{\partial x_k} x_j x_k + \ldots \qquad \text{[eqn 1.3]}$$

This can be repeated for all m defects. If we express this in vector/matrix notation then:

$$\boldsymbol{f} = \boldsymbol{f_o} + \boldsymbol{A}.\boldsymbol{x} + \ldots \qquad \text{[eqn 1.4]}$$

where **A** is the m x n matrix of the partial first differentials. If we substitute eqn 1.3, into the definition of the merit function, eqn 1.2, we find that:

$$\varphi = \sum_{i=1}^{m} f_{oi}^2 + 2\sum_{i=1}^{m} f_{oi} \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial f_i}{\partial x_j} \frac{\partial f_i}{\partial x_k} x_j x_k + \ldots \qquad \text{[eqn 1.5]}$$

The gradient, **G**, of the merit function, a n dimensional vector, is defined by the components G_i, where G_i, is the first derivative of the above with respect to the i^{th} variable, x_i.

$$G_i = 2 f_{oi} \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} + .. \qquad \text{[eqn 1.6]}$$

Typically, differentials are established by finite analysis, and are therefore time consuming to evaluate! This particularly applies to higher order derivatives, because of the huge numbers of them! However, if we are working near the starting point, or the defects are linear functions of the variables, then all higher order terms can be ignored. In this case, we can see by inspection, that:

$$\boldsymbol{G_o} = \boldsymbol{A}^T.\boldsymbol{f} \qquad \text{[eqn 1.7]}$$

Where $\mathbf{A^T}$ is the n x m transpose of **A**.

12.1.1 Steepest descent

Perhaps the most obvious method is that of the steepest descent, i.e. moving as in the direction of the gradient [steepest slope] at the starting point. We start off moving in the direction that a ball would follow if it were released on a slope. However, we either:

- Take small steps, evaluating the gradient at each point and making small adjustments to the direction at each step.
- Take larger steps, typically until the merit function starts to rise again. However, at this point the old gradient is now tangential to the merit function contours and the new gradient vector is therefore perpendicular to the old. Movement thus proceeds in a series of inefficient zigzag steps.

In either case many, many evaluations of the gradient are required. The solution will be reached, but only after a long, long time is past!

12.1.2 Least squares

As we have seen, the steepest descent method, though obvious and certain, has huge drawbacks!

Let us go back and reconsider our strategy. Ideally, we would be able to find a system in which all merit function contributions are reduced to zero. In practise, this is rarely achievable, so we will settle for minimising the merit function, i.e. minimising the residuals of the squares of the defects [hence 'least squares']

We will start by assuming that all defects are linear functions of the variables. In this case all higher order terms are identically zero, so:

$$\boldsymbol{f} = \boldsymbol{f_o} + \boldsymbol{A.x} \qquad \text{[eqn 1.8]}$$

and the merit function therefore becomes

$$\varphi = |\boldsymbol{f_o} + \boldsymbol{A.x}|^2 \qquad \text{[eqn 1.9]}$$

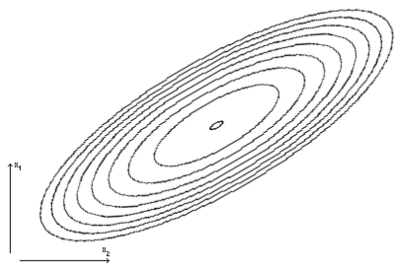
After expansion and some matrix manipulation, we obtain the result that:

$$\varphi = \boldsymbol{f_o}\,\boldsymbol{f_o} + 2\,\boldsymbol{A}^T\boldsymbol{f_o} + \mathbf{x}.\,\boldsymbol{A}^T\mathbf{Ax} \qquad \text{[eqn 1.10]}$$

and the gradient vector is

$$\boldsymbol{G} = 2\,\boldsymbol{A}^T\boldsymbol{f_o} + \boldsymbol{A}^T\mathbf{Ax} \quad \text{or} \quad \boldsymbol{G} = \boldsymbol{G_o} + \boldsymbol{A}^T\mathbf{Ax} \qquad \text{[eqn 1.11]}$$

Furthermore, where the defects are truly linear, then the merit function contours are 'concentric' hyper-ellipsoids[15].



Now, at the minimum of the merit function, the gradient or derivative of φ with respect to **x** is by definition, 0, i.e.

$$\mathbf{2}\,\boldsymbol{A}^T\boldsymbol{f_o} = -\,\boldsymbol{A}^T\mathbf{Ax} \qquad \text{[eqn 1.12]}$$

Then by a 'simple' matrix inversion we obtain the solution vector x.

Note that a single step takes us right from the starting point to the minima, compared to the thousands of steps required by the steepest descent method.

This may seem to be too good to be true, and of course its is. There are two problems:

- Linearly dependant variables
- Non-linear defects

If two or more variables are linearly dependent [or nearly so], then the ellipsoids become extremely extended and the matrix becomes ill conditioned. Small rounding errors become significant and lead to large errors in the solution vector. Of course, this can be overcome to an extent by using more and more precision or dropping one of a pair of such variables from the optimisation.

However non-linearity in the defects cannot be overcome by either method. Small amounts of non-linearity mean that the real minimum is not quite at the location of the least squares solution, and so several iterations will be required. Large non-linearity can mean that the LS solution is actually worse than the start point, and subsequent iterations will make matters even worse [divergent solution].

---

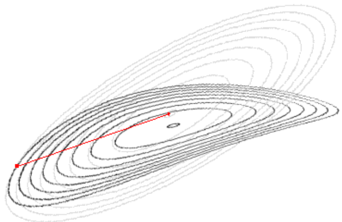
[15] Although perhaps not obvious from eqn 1.11, if we look back at eqn 1.5, then ignoring any higher order differentials, this can be re-written as:

$$\varphi = K + \sum_i^n B.xi + \sum_i^n C.x_i^2 + \sum_{i=1}^n \sum_{j=1}^{i-1} D.x_i x_j$$

This defines contours of ellipsoidal shape.

12.1.3 Damped least squares

We have seen the shortcomings of the least squares method in the presence of non-linearity in the defects. What is required is some method of constraining the solution vector [i.e. size of the changes in the variables].



This may be achieved by adding another term to the merit function, which depends upon the size of the solution.

$$\varphi = |\boldsymbol{f_o} + \boldsymbol{A.x}|^2 + |p\boldsymbol{x}|^2 \quad \text{[eqn 1.13]}$$

When eqn 13 is solved, we find the solution is given by:

$$\mathbf{2\ A^T f_o = -\ (A^T A + p^2 I)x'} \quad \text{[eqn 1.14]}$$

If p is zero, there is no damping and this reduces to the least squares solution.

If p -> infinity, then the solution vector is parallel to the gradient, $\mathbf{G_0}$, at the starting point, i.e. same direction as the steepest descent method. The actual solution magnitude is, of course, tiny.

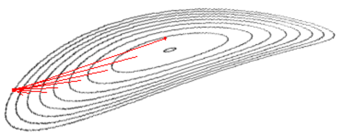
Typically p is somewhere in between. It is important to note that choosing different values of p do not merely scale the 'optimum' solution, but actually define different directions in variable space.

It is also important to understand that the p does not take non-linearity into account when predicting the solution, it merely aims to limit the size of the solution so that the non-linearity is not significant.

In summary, the effect of adding this extra term is to limit the size of the solution; hence p is usually known as the damping factor, because it tends to damp out massive changes. This also explains why this form of optimisation is known as damped least squares.

When damping is used, there is no possibility of reaching a minimum in one step [even if there is no non-linearity], and so several iterations will have to be used.

We now have to consider the obvious question: '**what is the 'best' damping factor?**'



Because of the unknown non-linearity, the 'best' damping factor cannot be predicted theoretically and has to be evaluated experimentally.

Moreover, the optimum damping factor will change from step to step of the optimisation.

The standard approach is outlined as follows:
1) Assess derivatives [sensitivities] **A** of defects at the starting point.
2) Calculate a solution **x** for each of several values of p
3) Evaluate the actual merit function for each solution **x**
4) Select the p, which leads to the best solution, i.e. smallest merit function.

The solution obtained in stage 4 becomes the starting point for the next cycle. Repeat stages 1 to 4, for as many cycles as required.

12.2 normalisation

A key part of the optimisation process, as implemented in WinLens, is normalisation.

First we note that our definition of the 'contribution' f, of each defect to the merit function, as shown below, is dimensionless:

$$f_i = (y_i - y^t_i) \,/\, tol_i \qquad \text{[eqn 2.1]}$$

This follows because the defect value, defect target and tolerance are all in the same units. All contributions can therefore be directly compared.

The second stage is the actual normalisation. The matrix **A** contains information on the partial derivatives of all defects in the merit function with respect to the variables.

Each column in this matrix contains the differentials for a single variable. Let us extract this column vector for the i[th] variable and denote it by $\mathbf{A}_i$. We will now normalise this variable change vector, by dividing it by its magnitude:

$$\boldsymbol{A}^*_i = \boldsymbol{A}_i \,/\, |\boldsymbol{A}_i| \qquad \text{[eqn 2.2]}$$

These then define a matrix, $\mathbf{A}^*$, of column normalised differentials.

In order to obtain the same contributions, as predicted by $\boldsymbol{f} = \boldsymbol{f_o} + \boldsymbol{A.x}$, it is also necessary to modify the variables:

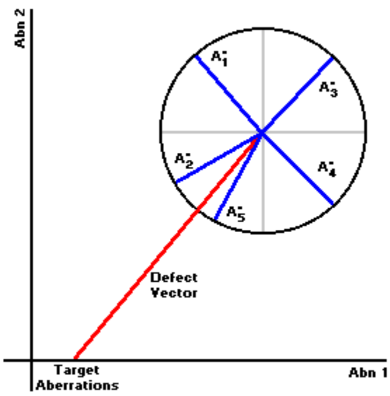
$$x^*_i = x_i \,.\, |\boldsymbol{A}_i| \qquad \text{[eqn 2.3]}$$

Thus we end up with a revision of eqn 1.8

$$\boldsymbol{f} = \boldsymbol{f_o} + \boldsymbol{A}^*.\boldsymbol{x}^* \qquad \text{[eqn 2.4]}$$

Optimisation proceeds with the values in the matrix **A***. The solution vector $\boldsymbol{x}^*$ is then re-transformed to obtain **x**.

The obvious question is why go through this extra stage? To understand why, let us plot the defect vector $\mathbf{f_o}$, and several variable vectors, $\mathbf{A_i}^*$.



We want to move from the starting system to the desired end solution, i.e. along the defect vector.

By simple inspection we can see that some of the variable vectors [1,3,5] are almost parallel or anti-parallel to the defect vector. Changing these parameters should then move us in the desired direction [ignoring non-linearity's], by altering the defects in the right proportions. Such variables are 'efficient'.

By contrast we see that others of these vectors [1,4] are almost perpendicular to the defect vector. No matter how large a change we apply, changing these variables will not help us achieve our target. Such variables are 'inefficient'.

Since all the variable vectors are normalised, a simple method of comparing their 'efficiency' is given by the angle between the defect vector and any given variable vector. In fact we use the variable efficiency [VE], where:

$$VE = \boldsymbol{f_o} \,.\, \boldsymbol{A_i} \,/\, |\boldsymbol{f_o}| \qquad \text{[eqn 2.5]}$$

The VE is derived from the scalar product of the two vectors, and is actually the cosine of the angle between them. These values are available to the user [see ***section 3.2.1.1***].

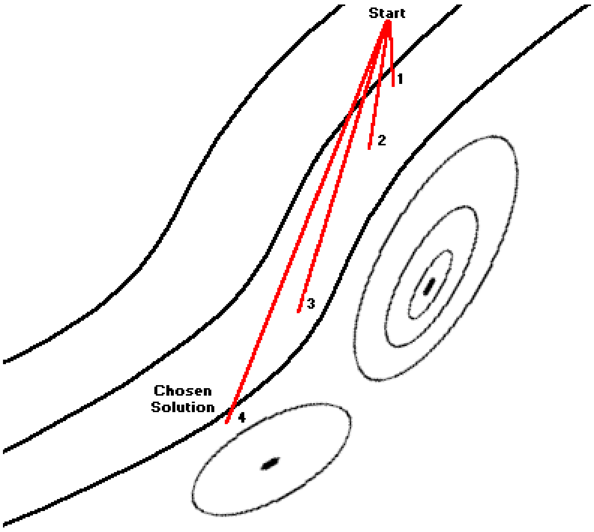
12.3 Line Search strategy

[***Section 5.1*** discusses how to activate this strategy]

When a solution for the DLS equations is obtained, both magnitude and 'direction' of that solution are determined by the damping factor, p.

Varying p changes the solution direction and magnitude in variable space.

Varying the p factor does not investigate the different magnitudes of solution along a single direction, i.e. does not look at the effect of scaling a solution. This can be a serious short coming. Consider the following illustration:

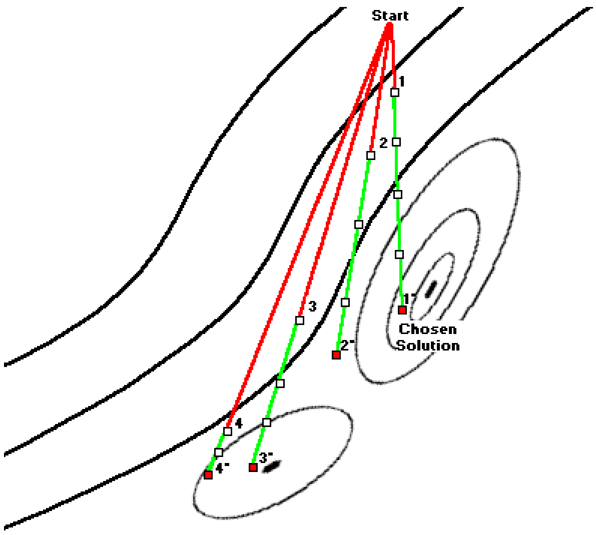


In this case the least damped solution [4] will be chosen by the standard method. The next cycle will then reach the nearest minima.

However it is clear that following this direction will never get to the better minimum!

In the line search strategy, we obtain, as before, the solution for a given damping factor. However, we also scale that solution several times, so we investigate further along the specified direction. This is repeated until we find the minimum value of the merit function along that particular direction.

This process is repeated for each damping factor, as shown below:



Finally we chose the solution with optimum damping and scaling.

In this case, the best solution is derived from the most highly damped case [1], but then scaled heavily [1*].

In this case the best solution is near the deeper minima and so will yield the better system

Of course one cycle of the line search strategy is not as fast as a single cycle of the standard method, but it does offer the possibility of reaching better solutions that would be missed by the standard method.

13 Index