

Saddle points in the merit function landscape of systems of thin lenses in contact

Florian Bociort^{*}, Alexander Serebriakov, Maarten van Turnhout

*Optics Research Group, Delft University of Technology
Lorentzweg 1, NL - 2628 CJ Delft, The Netherlands*

ABSTRACT

The merit function landscape of systems of thin lenses in contact, which are perhaps the simplest possible types of optical systems, shows remarkable regularities. It is easier to understand how the optimization parameter space of these simple systems is divided into basins of attraction for the various local minima if one focuses on the (Morse index 1) saddle points in the landscape rather than on the local minima themselves. The existence and the basic properties of these saddle points can be predicted by thin-lens theory, which is applied on a simplified model of the merit function containing only third-order spherical aberration. The predictions of this simplified model are confirmed by numerical results obtained with a typical merit function based on ray tracing.

Keywords: global optimization, saddle points, thin lens theory, spherical aberration, optical system design

1. INTRODUCTION

We have shown recently that, when certain quite general conditions are satisfied, the merit function landscape in optical system design has a remarkable property. The local minima form then a network in which all nodes are connected via links that contain a special type of saddle point¹. Saddle points are new concepts in optical system design. In order to understand them better we will analyze in this paper the perhaps simplest possible optical systems: lens systems consisting of spherical surfaces in which the distances between surfaces are negligibly small. We will show that the merit function landscape of these systems has surprising regularities, which can be modeled by thin-lens theory.

Figure 1 shows a typical saddle point in a merit function landscape with only two variables. As in the case of local minima, the gradient of the merit function vanishes at saddle points. (At a non-degenerate saddle point, the Hessian of the merit function, i.e. the matrix of the second-order derivatives of the merit function with respect to the optimization variables, has a nonzero determinant.) It can be seen that such a two-dimensional saddle point is a maximum in one direction (the downward direction), and a minimum along a direction (the upward direction) which is perpendicular on the first one. If we have $N > 2$ variables, some directions at the saddle point are upward, and some are downward directions. An important characteristic of saddle points is therefore the so-called Morse index, which is given by the number of downward directions. (Technically speaking, the Morse, or Hessian, index is equal to the number of negative eigenvalues of the Hessian; the "directions" are in fact the eigenvectors of the Hessian at the saddle point.)

For studying the network structure of the set of local minima, it is sufficient to consider the saddle points having a Morse index equal to 1 (SPMI1). Thus, a SPMI1 is a maximum in one direction (the downward direction), and a minimum in a $N-1$ -dimensional hyperplane orthogonal to that direction. Intuitively, the downward direction of a SPMI1 is similar to the downward direction in Fig.1, and each of the $N-1$ upward directions is similar to the upward direction in Fig.1. If we start local optimizations at two points situated on the downward direction close to the saddle point and on opposite sides of it, then local optimization will lead to two distinct local minima of the optical merit function

^{*} E-mail: bociort@tn.tudelft.nl, Fax: +31 15 2788105

landscape. Together with the SPMI1, the path in the parameter space generated by the two local optimizations forms a link between the corresponding local minima. In this way, the local minima in the optical merit function landscape are all linked via paths containing each a SPMI1¹.

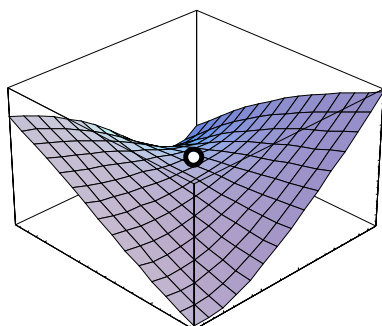


Fig. 1. In two dimensions every non-degenerate saddle point is a SPMI1.

A strategy described earlier^{2,3} to detect link by link the entire network of local minima has been implemented in our computer program NETMIN which uses for local optimization the commercial program CODE V. (The two programs communicate via the Windows COM interface.) A typical example of a network is shown in Fig. 2. Note that a two-dimensional graph captures the essentials of the topography in a merit function landscape having a high dimensionality. The lines between systems show how the systems are linked together via "minimum elevation" paths. For each such path, the merit function difference between the saddle point and the local minimum gives the height of the merit function barrier that must be overcome in order to reach from a given local minimum the basin of attraction of a neighboring one. If desired, the regions of interest in the network can be studied link by link, for instance by examining the way the optimization variables change along the links. In this way, additional insight in the highly complex merit function topography can be obtained.

When control parameters such as aperture, field or transverse magnification are varied, large parts of the network structure turn out to be robust, i.e. the local minima and saddle points change gradually, but their number and the connections between them tend to remain the same within large domains of variation of the control parameters. Other, more ephemeral, local minima and saddle points tend to appear or disappear more easily. This happens when the control parameters reach certain critical values.

In this paper, the emphasis will be on saddle points, because if these points are known, the local minima in the network result automatically. We will study the properties of the robust SPMI1 in the case of lens systems consisting of spherical surfaces in which the distances between surfaces are negligibly small. As will be seen in Secs. 3 and 4, these systems are simple enough so that basic properties of the set of saddle points can be predicted analytically. For simplicity, we begin in Section 2 with a two-dimensional analysis of a thin doublet.

2. TWO - DIMENSIONAL ANALYSIS OF A DOUBLET

Remarkable properties of the merit function landscape of systems of thin lenses in contact can already be observed for doublets if two-dimensional plots of the equimagnitude contours (i.e. contours along which the merit function is constant) are carefully analyzed. In Fig. 3 such a plot is shown for a doublet which, for simplicity, is monochromatic. (Axial color correction will be included for systems discussed in the next section.) As in the other cases discussed in the remainder of this paper, the object is at infinity and the merit function (MF) is the default error function of CODE V, for which the image defects are transverse ray aberrations computed with respect to the chief ray. If we use the curvature c_4 of the last surface to keep the focal length constant and use the curvature c_1 of the first surface as a control parameter, the two remaining curvatures can be used as variables. (In all plots in this section the curvature c_2 of the second surface is along the vertical axis and c_3 of the third surface is along the horizontal axis.) A plot very similar to Fig.3 has already been published by Sturlesi and O'Shea⁴. (See Fig 7. in their paper.)

In Fig. 3 we see four MF local minima and three MF saddle points. As expected, the best three local minima are located almost on the contours for third-order spherical aberration (SA) equal to zero and close to the zero coma contours. These local minima are also obtained by optimizing downwards from the three MF saddle points. In two-dimensional plots, the saddle points can be easily recognized. Simple mathematics shows that, close to the saddle point, the equimagnitude contours must be hyperbolas. For the value of the merit function corresponding to the saddle point itself the contours degenerate into a pair of straight lines (the asymptotes of the hyperbolas).

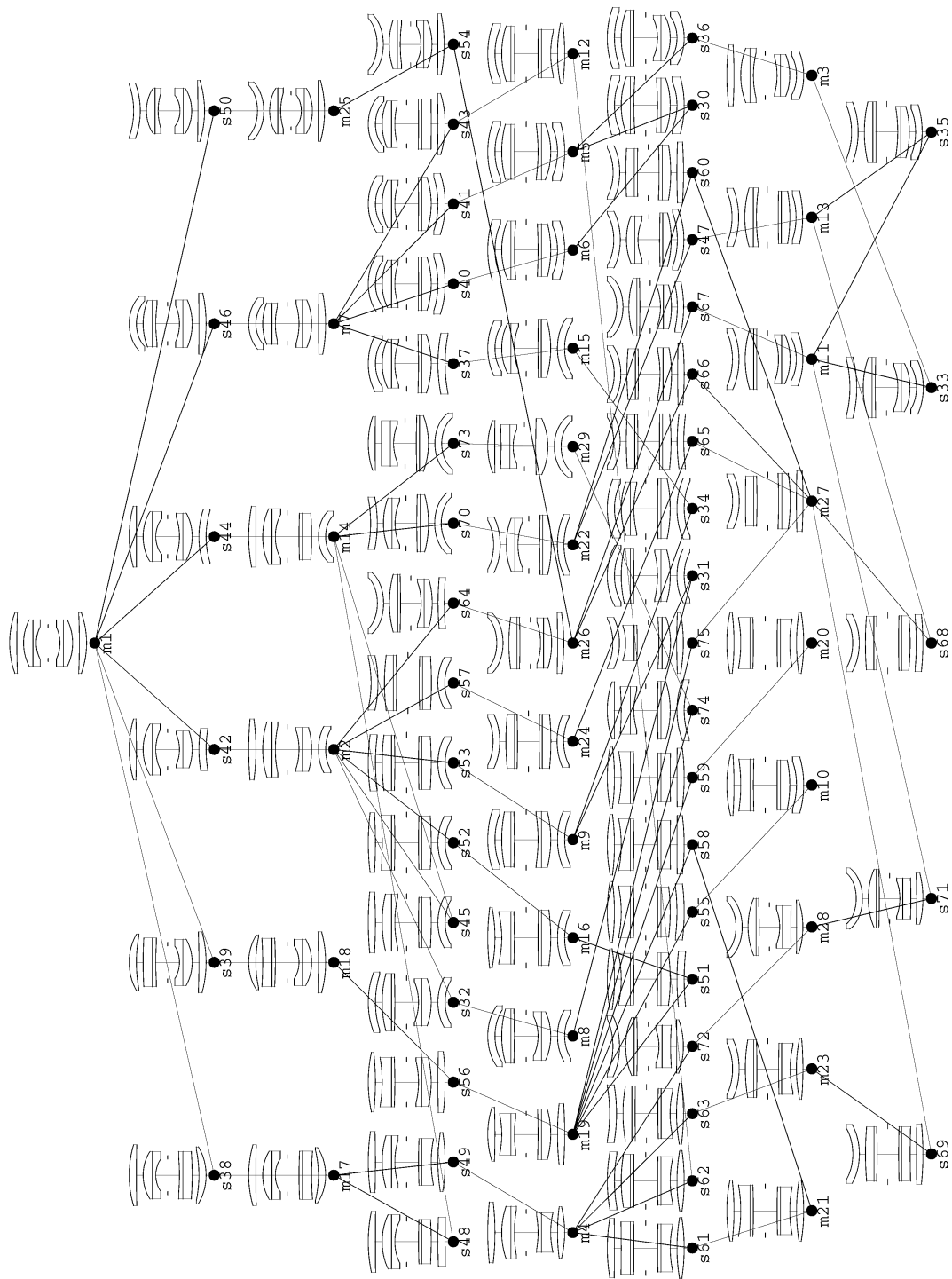


Figure 2. The network of local minima corresponding to a Double Gauss global search, as detected by the present version of the program NETMIN. The optimization variables are the curvatures of the surfaces, and the merit function is the default error function of CODE V. The "m" systems are local minima, and the "s" systems are saddle points with Morse index 1.

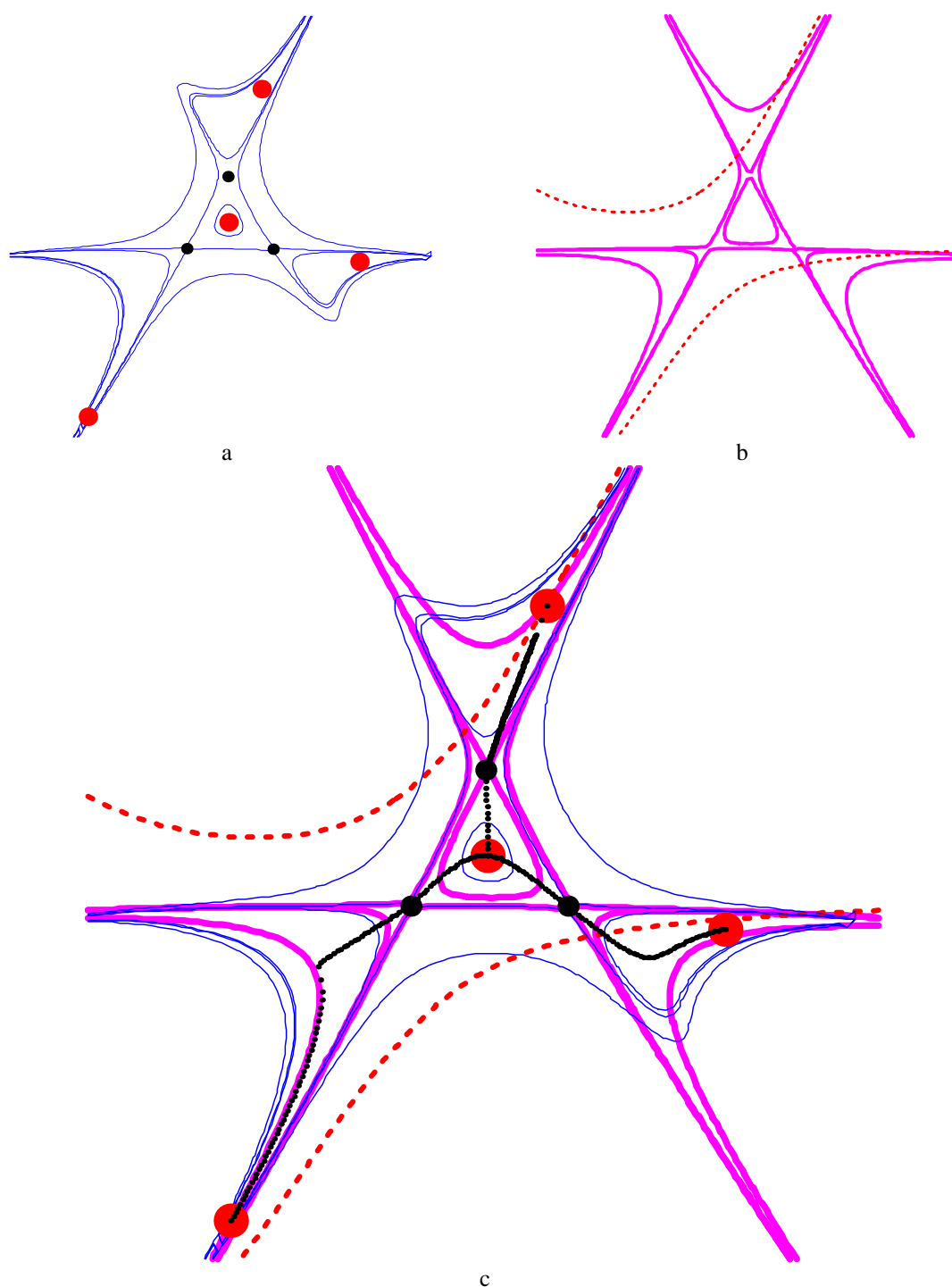


Figure 3. Topography for typical values of c_1 of a two-dimensional monochromatic doublet landscape (F number 3, field of 3 degrees, and the lens thicknesses are small). a) MF equimagnitude contours, local minima (large gray points) and saddle points (small black points). b) The contours for zero SA and the contours that correspond to the saddle points (thick gray lines). The contours for third-order coma equal to zero are shown as well (thin dashed lines). All data are computed with CODE V. Note how the saddle point SA contours separate the zero SA contours (the strongly curved ones) in three different branches. c) Figures a) and b) superimposed. In addition, the network links between saddle points and local minima are shown (dotted curves).

Note that the MF saddle points (the small black points in Fig.3) are situated very close to the SA saddle points (the points where the thick gray lines cross in Fig. 3b). A first remarkable property of the saddle points is that they are situated at the intersection points of three almost straight SA equimagnitude contours. In order to explain this and other surprising properties of the MF SPMI1, we will use in this paper a simplified merit function model in which

- i) only SA, the dominant aberration for the systems under investigation, is considered,
- ii) the imaging is monochromatic and all glasses are identical,
- iii) thin-lens theory will be used, i.e. in all formulas the distances between lens surfaces will be put equal to zero.

If in a power series expansion for the transverse aberration of a given ray we keep only the SA contribution and neglect the rest, then for all rays the transverse aberration will be proportional with SA. If MF is a weighted sum of squares for the transverse aberrations of a given set of rays, it will be proportional with SA^2 . Assuming that for MF SPMI1 the SA is nonzero, in our model these points, which are solutions of the system of equations $\nabla MF=0$, will also have the property $\nabla SA=0$. (The nabla operator ∇ has as components the partial derivatives with respect to the variables.)

In the thin-lens approximation, we can write the third-order spherical aberration of a doublet as

$$SA = h^4(f_0 + \alpha f_1 f_2 f_3) \quad (1)$$

where we have used the abbreviations

$$f_0 = \frac{n^2}{(n-1)^2} + \frac{c_1^2(n+2)}{n} - \frac{c_1(2n+1)}{n-1} \quad (2)$$

$$f_1 = c_1 - c_2 - \frac{1}{n-1} \quad (3)$$

$$f_2 = c_2 - c_3 \quad (4)$$

$$f_3 = 2c_2(n^2 - 1) + c_3(n+2) - \frac{n(2n+1)}{n-1} + c_1(n - 2n^2 + 4) \quad (5)$$

and

$$\alpha = \frac{n-1}{n} \quad (6)$$

The fourth curvature, which does not appear explicitly in Eqs. (1-6), is used to keep the total power equal to 1. In these formulas n is the refractive index of the two glasses and h is the marginal ray height at the lens. (Since the aperture stop is placed at the lens, h is equal to half the stop diameter.) One way of obtaining these equations is by using for the doublet the well-known Coddington thin-lens formulas⁵. Alternatively, these equations (and especially their equivalents for a larger number of lenses, which will be discussed in Sec.4) can be directly obtained with computer algebra software by putting the distances between surfaces equal to zero in the well-known formulas that give the surface by surface contributions to SA for lenses with finite thickness and rearranging the result. Also, these equations can be directly verified by computing the SA with commercial optical design software (again by putting the distances between surfaces equal to zero). For instance, CODE V lists the quantity $SA/(-2h)$.

The condition $\nabla SA=0$ with $\nabla = (\partial/\partial c_2, \partial/\partial c_3)$ leads to a system of two nonlinear equations for the two unknowns c_2 and c_3 . (Recall that c_1 is used in this section as a control parameter, not as a variable.) It follows after some simple algebra that this system of equations has four solutions. An examination of the eigenvalues of the Hessian at the solutions shows that three of them are saddle points. It turns out that the three saddle points are the solutions of the linear systems of equations $f_1=0$ and $f_2=0$, $f_1=0$ and $f_3=0$, $f_2=0$ and $f_3=0$, respectively. (Note that f_1 , f_2 , and f_3 are linear functions of the curvatures.)

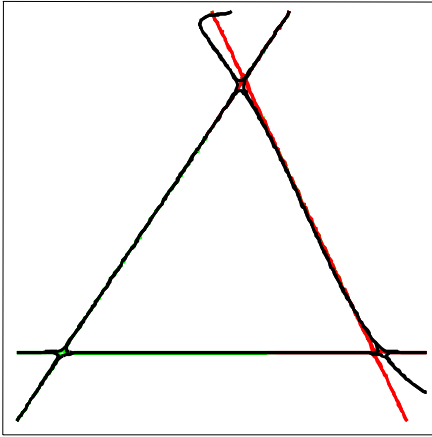


Figure 4. MF and SA equimagnitude contours that pass through the saddle points (MF drawn in black, SA in gray) for a doublet with zero thickness, $n=1.5$, F number 3 and field of 3 degrees.

Thus, in the two-dimensional SA plots the saddle points are located at the three intersections of the straight lines $f_1=0$, $f_2=0$, and $f_3=0$, taken in pairs of two. It follows from Eq. (1) that within the frame of our approximate model, along these lines SA remains unchanged, $SA = h^4 f_0$, i.e. these lines form an equimagnitude contour. In fact, the existence of two of these straight equimagnitude lines can be understood from basic principles even without using Eqs. (1-6). It can be easily shown from the fact that the fourth curvature is used to keep the total power equal to 1, that the condition $f_1=0$ means that along this line we have $c_3 = c_4$. The condition $f_2=0$ means $c_2 = c_3$. Since the distances between the surfaces are zero, in both cases all light rays pass undeviated through the corresponding pair of surfaces, regardless of the common value $c_3 = c_4$ or $c_2 = c_3$. Therefore, for zero thickness $f_1=0$ (horizontal line in Fig.4) and $f_2=0$ (left oblique line), remain equimagnitude contours even if other aberrations, including higher order ones, are included, i.e. these lines are MF contour lines as well. On the other hand the right oblique line is straight only for SA ($f_3=0$). Note that in Fig.3 the SA contour lines passing through the saddle points are slightly curved because of small but nonzero lens thickness.

In the middle of the triangle formed by the three saddle points we find a fourth solution with $\nabla SA=0$. (See Fig.3.) For values of the control parameter c_1 that are not too large, this point (which will be called in what follows the "hub" of the network) is a minimum. Despite the fact that merit function of the hub is poor, this minimum plays an essential role for the connectivity of the network. On one side, all three saddle points are linked with the hub.

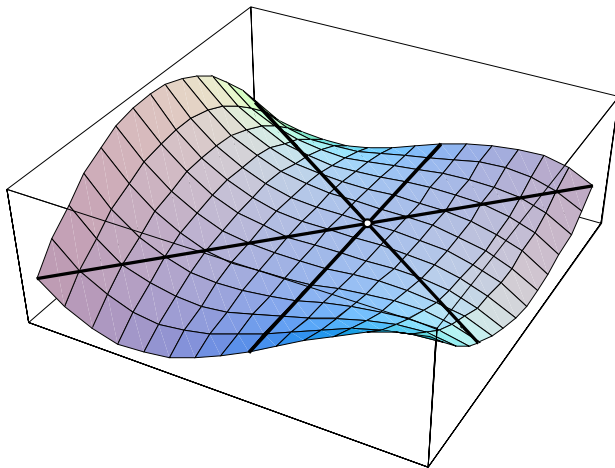


Figure 5. The SA "monkey saddle" point and the three straight equimagnitude lines that pass through it. At this point all elements of the Hessian matrix are zero, and the shape of the "monkey saddle" is determined by cubic terms.

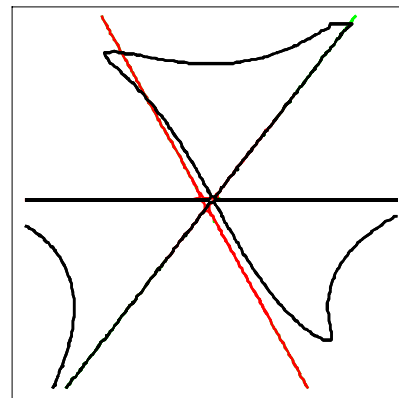


Figure 6. MF "monkey saddle" (black lines) for $c_{1,crit}=3.55$ in the same doublet as in Fig.4. Because the MF critical value is slightly different from the value given by Eq.(7), the SA line $f_3=0$ (gray) does not pass exactly through the MF monkey saddle.

When c_1 increases, the distances between the saddle points decrease and for values of c_1 larger than a critical value the hub becomes a maximum. (The network is then linked in a way that is different from the one shown in Fig.3.) Within the frame of the simplified model the critical value is given by

$$c_{1,crit} = \frac{n(2n+1)}{(n-1)(n+2)} \quad (7)$$

For example, for $n=1.5$ we have $c_{1,crit} = 24/7 = 3.43$. When $c_1 = c_{1,crit}$ the three equimagnitude lines pass through the same point. The three saddle points and the hub merge then into that single point. The SA landscape around this point takes then the peculiar shape sometimes called a "monkey saddle"⁶. (See Fig. 5. Note the difference between this figure and Fig.1.) For doublets with zero thickness, for a slightly shifted critical value of c_1 the "monkey saddle" can also be observed for the MF, where aberrations other than SA play a role as well (Fig.6).

3. NETWORK OF LOCAL MINIMA AND SADDLE POINTS FOR THE DOUBLET

For a monochromatic split doublet search, the network of MF local minima and SPMI1 detected with the program NETMIN is shown in Fig.7. The systems have small equal distances between surfaces and all three independent curvatures are used as variables. In the three-dimensional variable space, the local minima LM1 and LM5 and the saddle point S5 between them are situated close to the same SA=0 surface branch and the MF differences between them turn out to be low. In the network, the pair of systems formed by saddle point S5 and local minimum LM1 is therefore less robust than the rest of the network. As mentioned in Sec.1, such points can appear or disappear more easily when specifications or the merit function type are changed. When a saddle point disappears, one of the two local minima linked to it always disappears as well. The local minimum and the saddle point situated on the other side of the pair which has disappeared will then be linked together, so that such appearances or disappearances do not affect the connectivity of the network. A typical example for this behavior can be found in the Cooke triplet global search².

The other systems in Fig.7 are robust. For instance, their shape and the links between them are not affected by minor changes in the merit function definition. As mentioned before, for determining the links of a saddle point we use two points situated close to it on both sides along the downward direction. We have converted the pairs of CODE V files for the saddle points to the ZEMAX optical design program. Optimizing the converted files with the RMS Spot Radius merit function in ZEMAX leads to local minima that strongly resemble the corresponding local minima in Fig.7. (For the interested readers the files for the systems in the network are made available in both formats via our web site⁷.)

The existence of the robust systems in Fig. 7 can be explained with our simplified SA model. If all three independent curvatures of the thin doublet are used as variables, the equation $\nabla SA=0$ with $\nabla = (\partial/\partial c_1, \partial/\partial c_2, \partial/\partial c_3)$ has as solutions four SA SPMI1 and a minimum, the hub, at the center of the tetrahedron formed by the SPMI1. In the three-dimensional variable space, for each SPMI1 the downward direction (i.e. the direction of the eigenvector of the Hessian that has a negative eigenvalue) is approximately given by the line that passes through the given saddle point and through the hub. Along this direction, the SPMI1 is a maximum, whereas in a plane perpendicular to it the SPMI1 is a minimum. All SA SPMI1 have the same SA value. Interestingly, the SA value is also constant along all edges of the tetrahedron having the four SA SPMI1 as vertices.

In the two-dimensional analysis of the previous section we have found three saddle points, one at the crossing of the lines $c_2 = c_3$ and $c_3 = c_4$ and two others, one on each of these lines. One of the two-dimensional saddle points has thus three equal curvatures $c_2 = c_3 = c_4$, and the other two have two equal curvatures, $c_2 = c_3$ and $c_3 = c_4$, respectively. With the first curvature now variable, the four SPMI1 follow the same pattern. For two of them three successive curvatures are equal and for the other two SPMI1 two curvatures are equal, the first two curvatures and the last two curvatures, respectively. (See Table 1.) Note that in Fig.7 the robust MF saddle points, which have been detected numerically, have these properties as well. As predicted by the simplified model, on one side all the robust saddle points are linked with

4. SYSTEMS OF MORE THAN TWO THIN LENSES IN CONTACT

For thin lenses in contact, the simplified zero-thickness SA model predicts that robust the SPMI1 in the corresponding networks form an amazingly regular pattern. For $m > 1$ lenses, $2m-1$ curvatures from the total set of $2m$ curvatures are independent variables. It turns out that there are $2m$ SA SPMI1 that form a hyper-tetrahedron having again a hub at its center. When the total power is 1, the sets of curvatures for the saddle points can be obtained as follows: Consider first the sequence of points p_1, p_2, p_3, \dots given by

$$p_{2k} = p_1 + (k-1)d, \quad p_{2k+1} = -p_1 + (k-1)d \quad (8)$$

with $k=1,2,3,\dots$ and

$$p_1 = -\frac{n(2n+1)}{2(n-1)(n+2)}, \quad d = \frac{2(n+1)}{n+2} \quad (9)$$

The first SA SPMI1, S_1 , has then the curvatures $c_{1i} = p_i/(m-1)$, with $i=1,\dots,2m$. For $k > 1$, the curvatures of the saddle point S_k are given by $c_{k,k-1} = c_{kk} = p_{k+1}/(m-1)$ and $c_{ki} = c_{k-1,i}$ for $i \neq k-1$ and $i \neq k$. Tables 1 and 2 show the resulting values for the curvatures of the SPMI1 in the case of a doublet and a triplet, respectively. Note that the 4×4 curvature matrix in Tab.1 differs from the corresponding 4×4 submatrix (shown with bold italic font) in Tab.2 only by a factor 2 (that results from the division with $m-1$).

In the general case, we encounter a structure that is a generalization of the one for the doublet. The first SA SPMI1 has $c_1 = c_2$, for $2m > k > 1$ S_k has $c_{k,k-1} = c_{kk} = c_{k,k+1}$, and S_{2m} has $c_{2m,2m-1} = c_{2m,2m}$.

	Equal curvatures	c_1	c_2	c_3	c_4	c_5	c_6
S1	$c_1 = c_2$	-6/7	-6/7	6/7	-1/7	11/7	4/7
S2	$c_1 = c_2 = c_3$	6/7	6/7	6/7	-1/7	11/7	4/7
S3	$c_2 = c_3 = c_4$	6/7	-1/7	-1/7	-1/7	11/7	4/7
S4	$c_3 = c_4 = c_5$	6/7	-1/7	11/7	11/7	11/7	4/7
S5	$c_4 = c_5 = c_6$	6/7	-1/7	11/7	4/7	4/7	4/7
S6	$c_5 = c_6$	6/7	-1/7	11/7	4/7	16/7	16/7

Table 2. The curvatures of the six SA SPMI1 for a thin-lens triplet with $n=1.5$

Figure 8 shows the robust MF SPMI1 and the hub linked to them for a monochromatic quartet with small, but nonzero distances between surfaces. Despite the fact that the values of the curvatures differ somewhat from the ones computed with the simplified SA model, the structure mentioned above remains valid for the MF saddle points detected numerically. Moreover, note that, when we move from a MF saddle point to the following one, only one pair of consecutive surfaces changes significantly, while the other surfaces remain basically unchanged, as predicted by the simplified model. For a triplet, a figure having the same properties as Fig.8 can be obtained as well.

While for a doublet the hub is a poor quality local minimum, Fulcher has shown that for $m=3$ and $m=4$ the system called in this paper a hub can have $SA=0$ for a certain value of the refractive index¹¹. For instance, the quartet hub has $SA=0$ for $n=1.5$. Moreover, the quartet hub is known to be a relaxed design that has an axial imaging of excellent quality even at large apertures^{8,12}.

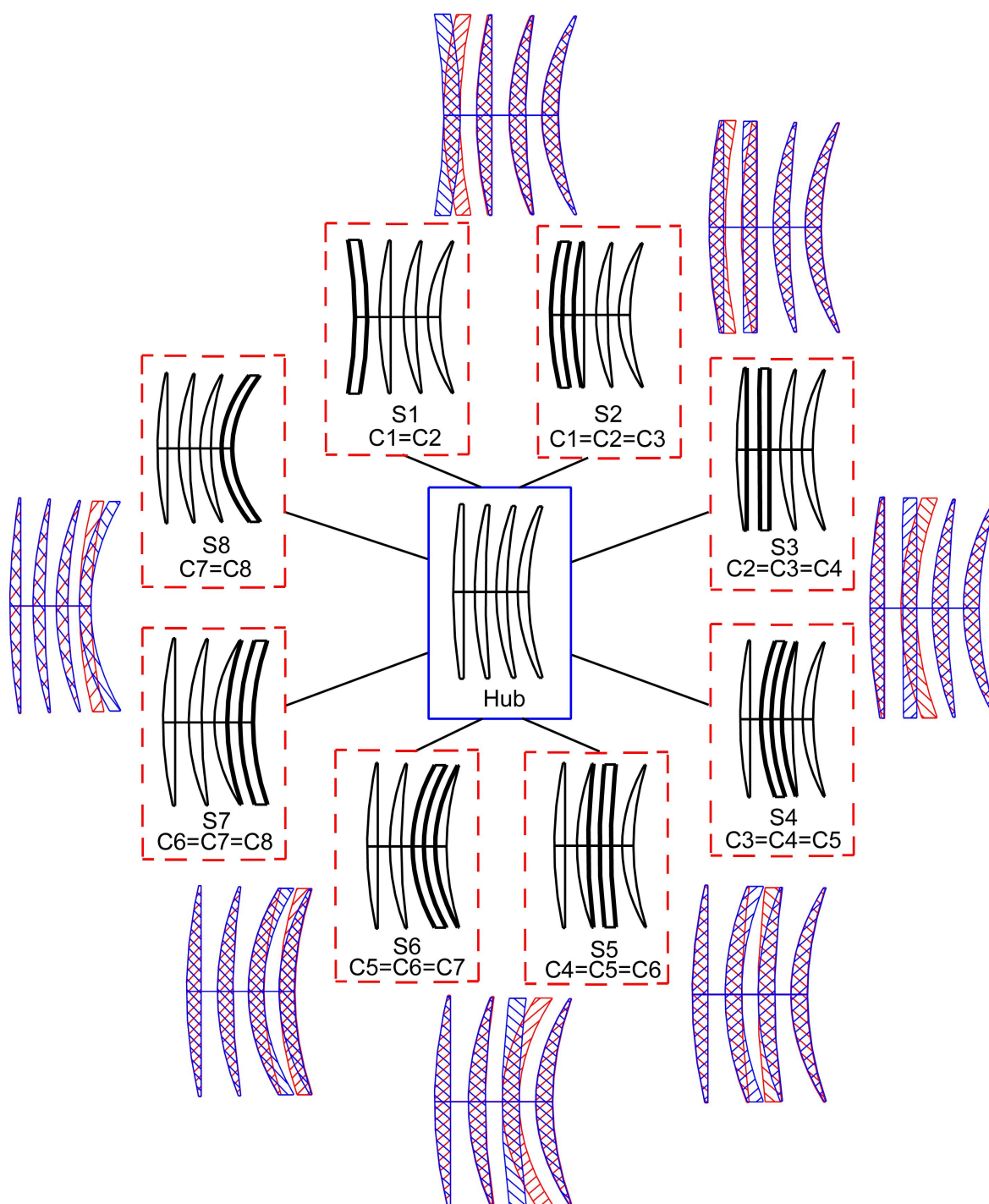


Fig. 8. The hub and the eight robust MF SPMI1 around it in the case of a monochromatic quartet topography. All systems have been detected numerically. The systems have F number 5, field 3 degrees, and $n=1.5$. The distances between consecutive surfaces are equal, such that the total distance between the first and the last surface is 14% of the focal length. For better comparison, on the outer ring figures of successive saddle points have been superimposed. Since the curvatures of the systems are small, the aspect ratio of the plots has been changed so that the curvatures appear larger. Since this graphical change causes the thicknesses to appear too large, the thicknesses have also been reduced in the drawings.

5. CONCLUSIONS

With present-day knowledge, one of the major difficulties in global optimization is that, if a given local minimum is known, it is extremely difficult or even impossible to have a basic understanding of where the basin of attraction of the given minimum ends and the basins of attraction of neighboring local minima begin. The various global optimization techniques find the new basins of attraction via techniques that involve a substantial amount of brute-force search. In this paper we show an exception to this rule. When some order exists in the merit-function landscape, it is more easily discovered if one focuses on the saddle points rather than on the local minima. Once the saddle points are known, the corresponding local minima result automatically by following the downward links from the saddle points. For systems of thin lenses in contact, a simplified model can predict analytically the most important saddle points in the merit function landscape. This analysis reveals the way in which the variable space is divided between the major basins of attraction.

ACKNOWLEDGMENTS

We would like to thank Joseph Braat for stimulating discussions and Eco van Driel for writing the computer program NETMIN. We also acknowledge the use of an educational license of CODE V.

REFERENCES

1. F. Bociort, E. van Driel, A. Serebriakov, "Networks of local minima in optical system optimization", *Optics Letters* **29**(2), p. 189-191 (2004)
2. F. Bociort, E. van Driel, A. Serebriakov, "Network structure of the set of local minima in optical system optimization", *Proc. SPIE* **5174**, p. 26-34 (2003)
3. E. van Driel, F. Bociort, and A. Serebriakov, "Topography of the merit function landscape in optical system design," *Proc. SPIE* **5249**, p. 353-363 (2004)
4. D. Sturlesi, D. O'Shea, "Global view of optical design space", *Opt. Eng.* **30**(2), p.207-217 (1991)
5. W.T. Welford, *Aberrations of optical systems*, Adam Hilger, Bristol (1986)
6. E. W. Weisstein, "Monkey Saddle" in *CRC concise encyclopedia of mathematics*, CRC Press, Boca Raton, (1999), (also available at Wolfram Web Resource <http://mathworld.wolfram.com/MonkeySaddle.html>)
7. http://wwwoptica.tn.tudelft.nl/users/bociort/doublet_network.html For more significant merit function changes, the two systems for a given saddle point must be farther away from the saddle point itself, so that these systems enter more deeply in the basins of attraction of the corresponding local minima.
8. D. Shafer, "Optical Design and the Relaxation Response", *Proc. SPIE* **0766**, p. 2-9 (1987)
9. W. J. Smith, *Modern optical engineering: the design of optical systems*, McGraw-Hill, New York (2000)
10. M. J. Kidger "Techniques for Handling Difficult Lens Design Problems 1. The 'Local Minimum' Problem", *Proc. SPIE* **0766**, p. 27-31 (1987)
11. G. S. Fulcher, "Telescope objective without spherical aberration for large apertures, consisting of four crown glass lenses", *JOSA* **37**(1), p. 47-53 (1947)
12. R. E. Fisher, "Spherical Aberration - some fascinating observations", *Proc. SPIE* **0766**, p. 53-60 (1987)