Why are there so many system shapes in lens design?

F. Bociort^{*} Delft University of Technology, Optics Research Group Lorentzweg 1, NL - 2628 CJ Delft, The Netherlands

ABSTRACT

The presence of many local minima in the merit function landscape is perhaps the most difficult challenge in lens design. We present a simplified mathematical model that illustrates why the number of local minima increases rapidly with each additional lens added to the imaging system. Comparisons with results obtained with lens design software are made for the design landscape of triplets with variable curvatures, a problem that is nontrivial, but still simple enough to be analyzed in detail. The mathematical model predicts how many types of local minima can exist in the landscape of the global optimization problem and what are, roughly, their curvatures. This model is mathematically quite general and might perhaps be useful as an analogy for understanding other global optimization problems as well, there where the number of local minima increases rapidly when more components of the same kind are added in the model of the problem.

Keywords: saddle point; critical point; global optimization; optical system design; complexity

1. INTRODUCTION

The presence of many local minima in the merit function landscape is perhaps the most difficult challenge in lens design. Refraction of rays at a surface is described by a simple law that makes the design of a single lens, which is optimal for a given application, rather straightforward. However, increasing the number of lenses in an imaging system makes optimal design much more complex, because the topological structure of the design landscape changes qualitatively: the number of local minima that are present increases rapidly with the number of lenses. A good choice of a starting configuration that leads after optimization to a design with high imaging quality is then essential. The theory of primary aberrations may be useful for this purpose if the system to be designed is sufficiently simple (e.g. an achromatic doublet) but for more complex systems struggling with the labyrinthine structure of a design space with many local minima consumes a significant part of the total design effort.

A fundamental approach for finding the best possible design is to answer two different questions: first, how many types of local minima can exist and what are, roughly, their constructional parameters and second, which local minimum in this set is the best one. In this work the first question will be addressed. It will be shown that the presence of many local minima in the design landscape is a fundamental problem, not one caused e.g. by imperfect design techniques. We will focus here on what we called earlier¹ system (or design) shapes. Local minima that are close to each other in the variable space and have almost the same merit function values correspond to the same system shape (see Ref. 1 for more details).

For a better understanding of the properties of the design landscape it is useful to organize the set of local minima that result from a global optimization run in networks^{2,3}. For a typical global optimization run such a network may seem rather complicated and irregular⁴. For global optimization runs for simple systems (doublets and triplets) with variable curvatures, we have observed that the complexity of the network structure results from the interplay of two different mechanisms, one that generates new local minima when new lenses are added, and another one that destroys local minima when specifications (e.g. aperture and field) or design parameters that are not optimization variables are changed. By a careful choice of settings, we can obtain a theoretically interesting situation where only the first mechanism is present. If we draw only one minimum per system shape in the network, we have observed that the corresponding network becomes perfectly regular (the so-called ideal or fundamental network)¹. The system shapes that appear in the fundamental network form a set having maximal diversity, i.e. in the practical runs we have studied some

Optical Design and Testing IV, edited by Yongtian Wang, Julie Bentley, Chunlei Du, Kimio Tatsuno, Hendrik P. Urbach, Proc. of SPIE Vol. 7849, 78490D · © 2010 SPIE · CCC code: 0277-786X/10/\$18 · doi: 10.1117/12.873880

^{*} email: F.Bociort@tudelft.nl

fundamental shapes may not be present, but no new system shape could be found. When, as e.g. in the case of practical settings, the second mechanism is in action, the structure of the fundamental network helps us to understand not only the structure of the more complicated real network, but also why certain fundamental design shapes are not present in the corresponding design landscape¹.

In Secs. 2-5 we present a simplified mathematical model derived from aberration theory that captures the essence of the topology of fundamental networks and explains, for simple optical systems, what types of local minima can exist in the merit function landscape. Examples will be given in Sec. 6 for the design landscape of triplets with curvatures as variables, for which the design landscape is already fairly complex, but still manageable. In the Appendix a method called Saddle-Point Construction, which was described earlier in the specific context of lens design⁵, will be discussed in a more general context, which is appropriate for the mathematical model used in Sec.2.

2. TOY MODEL

A first goal of the present work is to develop the simplest possible mathematical model that explains the qualitative properties (e.g. the number of system shapes, the perfectly regular structure) of the fundamental network for three lenses (shown in Figure 3 of Ref. 1) and that predicts, roughly, the constructional parameters of all systems that appear there. The global optimization problem there has N=6 variables (the six surface curvatures) and there is one (paraxial) constraint (e.g. effective focal length constant or total track constant, see Ref. 1 for more details). A second, even more important goal, will be to use this model (which will be called the "toy" model, because it is simplified to the extreme) in order to illustrate why the number of local minima increases rapidly with each additional lens added to the imaging system. Finally, the toy model is mathematically so general that it might be useful as an analogy that helps understanding other global optimization problems as well, there where the number of local minima increases rapidly when more components of the same kind are added in the design model. Because of the fundamental nature of the problems discussed here, they are best understood when the treatment is kept at a general mathematical level. Therefore, the specific optical details are intentionally postponed for Sec.6.

As mentioned earlier⁶, and as will be discussed in more detail in a subsequent paper, the starting point for the derivation of the toy model is thin-lens aberration theory. Since only qualitative properties are envisaged, gross approximations can be made, e.g. not even the refractive index of the lens material is explicitly present in the model. The refractive index will appear in the formulas only in the final stage, when the *N* variables $(z_1, z_2, ..., z_N)$ are translated into lens surface curvatures using the formulas given in Ref. 6.

In the toy model, the simplified "merit" function is given by

$$f(z_1, z_2, \dots z_N) = \left(\sum_{k=1}^N z_k^3\right)^2 + a \left(\sum_{k=1}^N z_k^5\right)^2$$
(1)

subject to the constraint

$$\sum_{k=1}^{N} z_k = t \tag{2}$$

When numerical values are required, in this work we will use a=1/100 and t=1.

Note the high degree of symmetry present in this model. For any point in the variable space, the value of f will not be affected by permutations of the variables (e.g. by interchanging z_i, z_k for arbitrary i and k).

For studying the "landscape" of the toy model, it is important to consider not only the local minima that are present, but other critical points as well. For handling critical points in the presence of the constraint (2) it is useful to use the Lagrange multiplier method⁷ because it preserves the symmetry of the toy model. We first compute the associated Lagrange function

$$f^{*}(z_{1}, z_{2}, \dots z_{N}, b) = \left(\sum_{k=1}^{N} z_{k}^{3}\right)^{2} + a \left(\sum_{k=1}^{N} z_{k}^{5}\right)^{2} + b \sum_{k=1}^{N} z_{k}$$
(3)

where b is the so-called Lagrange multiplier, and the partial derivatives

$$\partial f^* / \partial z_k = b + 6z_k^2 \left(\sum_{k=1}^N z_k^3 \right) + 10az_k^4 \left(\sum_{k=1}^N z_k^5 \right)$$
(4)

A point in the landscape is a critical point if a value of b can be found such that for all k

$$\partial f^* / \partial z_k = 0 \tag{5}$$

The constraint (2) must also be satisfied.

An important characteristic of a (non-degenerate) critical point is its so-called Morse (or Hessian) Index. Critical points with Morse Index 0 are local minima, those with Morse Index N are maxima, and those with a Morse Index between 1 and N-1 are called saddle points (for more details see e.g. Ref. 6). For computing the Morse index, we eliminate e.g. the last variable by using the constraint equation (2)

$$z_N = t - \sum_{k=1}^{N-1} z_k$$
(6)

The function obtained by substituting Eq.(6) in Eq.(1) will be denoted by $\tilde{f}(z_1, z_2, ..., z_{N-1})$. We then compute the

Hessian matrix having the elements $H_{ij} = \frac{\partial^2 \tilde{f}}{\partial z_i \partial z_j}$ with i,j=1...N-1. To obtain the Morse Index, we insert in the Hessian

the set of the first N-1 z values of the critical point and compute with standard numerical techniques the eigenvalues of the resulting matrix. The number of negative eigenvalues is then the Morse Index of the critical point. When one or more eigenvalues are zero, the critical point is degenerate.

If we want to find the critical points of the toy model, solving the system of equations (4-5) directly is very difficult, because of the high-order polynomials appearing there. Therefore we will use two heuristic methods to find the critical points that are of interest for the present purpose. If a heuristic method suggests that a point is a critical point, this can be easily verified by solving for *b* e.g. the equation $\partial f^* / \partial z_1 = 0$ and then by checking whether for the same value of *b* all other equations (5) are satisfied. For the first of the two heuristic methods, we will use the concept of the so-called fundamental critical point.

3. FUNDAMENTAL CRITICAL POINTS

We define the fundamental critical points as the critical points of the toy model that are also critical points of the function

$$T = \sum_{k=1}^{N} z_k^{3}$$
(7)

subject to the constraint (2).

We want now to find all critical points of Eq.(7). To transform this constrained problem into an unconstrained one we use again the Lagrange multiplier method. With μ as Lagrange multiplier, we now look for the critical points of the Lagrange function associated to T

$$T^* = \sum_{k=1}^{N} z_k^{3} - \mu \sum_{k=1}^{N} z_k$$
(8)

From the equations

$$\partial T^* / \partial z_k = 3z_k^2 - \mu = 0 \tag{9}$$

it can be immediately seen that for a given critical point the absolute value of the variables is the same for all k

$$\left|z_{k}\right| = \sqrt{\mu/3} = \omega \tag{10}$$

and we can write

$$z_k = \pm \omega \tag{11}$$

The set of critical points is obtained by permuting the plus and minus signs in all possible ways, and ω is chosen so that the constraint (2) is satisfied. The critical points can then be classified according to the number of negative signs (denoted by *MI*) in the solution vector $z = (z_1, z_2, ..., z_N)$. We therefore have

$$\omega_{MI} = t / (N - 2MI) \tag{12}$$

It can be easily seen that a critical point of the function (7) is a fundamental critical point, i.e. it is also a critical point of Eq.(1) independent of the value of a. Because of Eq.(11), the sum S_3 of cubes and the sum S_5 of fifth powers are the same for all critical points of Eq.(7) with the same MI

$$S_{3} = \sum_{k=1}^{N} z_{k}^{3} = (N - 2MI) \omega_{MI}^{3}$$

$$S_{5} = \sum_{k=1}^{N} z_{k}^{5} = (N - 2MI) \omega_{MI}^{5}$$
(13)

Since $z_k = \pm \omega$, for all k, for

$$b = -6\omega_{MI}^{2}S_{3} - 10a\omega_{MI}^{4}S_{5}$$
(14)

we have $\partial f^* / \partial z_k = 0$ for all k variables of the given critical point. By computing the eigenvalues of the Hessian as shown above, it turns out that the number *MI* of minus signs in the solution vector of a given fundamental critical point is exactly its Morse Index.

As will be seen in what follows, a special role in the present analysis is played by the critical point with MI = 0 having

$$z_k = t / N \tag{15}$$

for all k. This local minimum that exists for all values of N will be called the "main hub" of the landscape. In Eq. (12) all values of MI < N/2 need to be considered. In this paper, the maximal value of N will be 6, which is necessary for explaining the triplet landscape. For N=2 (singlet lens) the only fundamental critical point is the main hub. When N increases, additional types of fundamental critical points appear. For N=3 and N=4 we also have saddle points with MI = I, and for N=5 and N=6 saddle points with MI = 2 appear. As shown below, the appearance of new types of fundamental critical points when N increases leads to a significant increase of the complexity of the design landscape.

4. CRITICAL-POINT PROJECTION

For simplicity, we consider first the case N=3. Because of the constraint (2) the analysis of the critical points is now a two-dimensional problem. In order to preserve the symmetry of the problem, rather than eliminating a variable by using Eq. (6), we rotate the coordinate system such that one of the new axes, t, is perpendicular on the constraint plane $z_1 + z_2 + z_3 = t$ and the other two, u_1, u_2 , are in the constraint plane. By inserting in Eq.(1)

$$z_{1} = t/3 + u_{1}/\sqrt{2} - u_{2}/\sqrt{6}$$

$$z_{2} = t/3 + \sqrt{2/3}u_{2}$$

$$z_{3} = t/3 - u_{1}/\sqrt{2} - u_{2}/\sqrt{6}$$
(16)

and also t=1 and a=0.01, f becomes a function of two independent variables u_1, u_2 . In this two-dimensional landscape we can find four local minima and three saddle points with MI = 1 as shown in Fig. 1a. The central local minimum (at $u_1 = u_2 = 0$) is the main hub, which according to Eq. (15) has z = (1/3, 1/3, 1/3). The three saddle points are fundamental critical points for which, according to Eqs. (11) and (12) we have z = (-1, 1, 1), z = (1, -1, 1) and z = (1, 1, -1), respectively. Note that the coordinate vectors for the last two saddle points are simple permutations of the coordinates of the first saddle point. On one side of any of the three saddle points, the optimization rolls down to the same minimum, the main hub. On the other side, each saddle point leads to a different local minimum. It is important to note that, because of the high symmetry, the local minima other than the main hub are situated exactly on the line that passes through the main hub and through one of the saddle points, that are fundamental critical points. Figure 1b shows that, on the side opposite to the main hub of the corresponding saddle point, all equimagnitute contours are perpendicular on any of these three lines. Therefore, the derivative along the direction perpendicular on these lines is zero. In order to find the z values for these three minima, it is sufficient to look for them along the lines passing through the main hub and saddle points. For instance, the line through the main hub and the saddle point with z = (-1, 1, 1) is given by

$$z_1 = 1/3 - 4u/3, \ z_2 = z_3 = 1/3 + 2u/3 \tag{17}$$

where the variable *u* gives the position along this line, e.g. u=0 gives the main hub and u=1 gives the saddle point with z = (-1, 1, 1). Inserting Eq. (17) in Eq.(1) leads to a function with a single variable *u* that has three extrema, as shown in Fig.2.



Fig.1 Equimagnitude contours of the toy model function for N=3, with u_1 on the horizontal axis, and u_2 on the vertical one. Local minima are shown with small dots, saddle points (situated there where two equimagnitute lines cross) are shown with large dots. For keeping Fig. 1a simple, only a few equimagnitute contours are shown there. In Fig. 1b the number of displayed equimagnitute contours is larger, in order to show that they are perpendicular on the line passing through the main hub and saddle points.



Fig.2 Typical behavior of the toy model function along a line passing through the main hub and a fundamental critical point, with *u* on the horizontal axis, and the value of the toy model function on the vertical one

With a=0.01, in addition to the extrema for u=0 (the main hub) and u=1 (the saddle point) we find a third extremum for u=1.46426. Inserting this value of u in Eq.(17) gives z = (-1.61902, 1.30951, 1.30951). By considering the possible permutations, the three peripheric minima shown in Fig. 1a have z = (-Y,y,y), z = (y,-Y,y), and z = (y,y,-Y) respectively, with Y=1.61902 and y=1.30951. By using Eqs. (4) and (5) it can be easily checked that these three points are indeed critical points for b = 1.53836.

A remarkable property of the toy model is that for N=3,4,5,6 all local minima, other than the main hub, which are relevant for the present research, can be obtained in the same way. As in the simple case shown in Fig. 1, there is a one-to-one correspondence between these local minima and the fundamental critical points resulting from Eqs.(11) and (12). For each fundamental critical point *FCP* other than the main hub, a local minimum can be found on the line

$$z_{k}(u) = 1/N + u(z_{k}(FCP) - 1/N) \quad \text{for } k=1...N$$
(18)

passing though the main hub given by Eq. (15) (here with t=1) and the given fundamental critical point. By inserting the line (18) into Eq. (1) we obtain a function with a single variable u that has a plot very similar to the one shown in Fig. 2. By inserting the u value of the minimum on the opposite side of the main hub (u>1) in Eq. (18) we obtain the coordinates of a local minimum that is called the projection of *FCP*. This heuristic method to obtain local minima is called critical-point projection.

For a given N it is sufficient to consider only one fundamental critical point for each allowed Morse Index value. Because of the high symmetry of the model, other fundamental critical points with the same Morse Index will lead to the same u value for their projected minimum.

For N=4 (which in lens design corresponds to the design landscape of the airspaced doublet), the fundamental critical points are the main hub given by Eq. (15) and four saddle points with Morse Index 1, given by

z = (-1/2, 1/2, 1/2, 1/2), z = (1/2, -1/2, 1/2, 1/2), z = (1/2, 1/2, -1/2, 1/2), z = (1/2, 1/2, 1/2, -1/2)

respectively. They lead to four projected local minima which for a = 0.01 have u = 1.56067. From Eq. (18) we obtain then the coordinates of the four projected minima

z = (-Y, y, y, y), z = (y, -Y, y, y), z = (y, y, -Y, y), z = (y, y, -Y, y), z = (y, y, y, -Y)respectively, with Y=0.9205 and y=0.640167.

The one-to-one correspondence observed in Fig. 1 between fundamental critical points and local minima is more general than the present toy model. We have observed it also in more elaborate models, which include terms that correspond to additional aberrations not present in Eqs. (1) and (2) (to be discussed elsewhere) and we have observed it in doublet and triplet global optimization landscapes obtained with commercial optical design software⁶. There however, the technique to obtain the local minima from the corresponding fundamental critical points is more elaborate, because the main hub, the fundamental critical points and their projected minima are not situated on straight lines any more.

We note also that in general critical-point projection cannot find all minima present in the landscape. For instance, in the present model (Eqs. (1) and (2)) we can find for N=6 points for which the two sums of third and fifth-order terms on the right-hand side of Eq.(1) vanish simultaneously. These points are degenerate critical points (i.e. critical points having zero eigenvalues). However, if additional terms are added in Eq. (1) e.g. a squared sum of 7th-order terms, the degeneracy is removed and those critical points become 60 new local minima. These new local minima cannot be found via critical-point projection, but can be found via local optimization by using the technique described in the next section and in the Appendix. These new local minima are in a sense less fundamental than the 22 local minima found earlier for N=6, because when translated to lens drawings, each of them strongly resembles one of the 15 minima that result from Morse index 2 projection, and cannot therefore be called new "system shapes". For each Morse index 2 projection, there are four such new minima similar to it (i.e. for any k, the z_k -values for all these five systems have the same sign).

5. SADDLE-POINT CONSTRUCTION AND NETWORK OF THE TOY MODEL

For understanding the properties of an optimization landscape it is useful to consider not only the local minima existing there, but also all saddle points with Morse index 1. For example, it can be observed in Fig. 1a that for each saddle point with Morse index 1, on one side of the saddle optimization rolls down to one minimum, on the other side it leads to a different minimum (the straight lines from the saddle points to the minima indicate the optimization paths). Saddle points with Morse index 1 have the same useful property even when N increases: on the two sides of such a saddle point optimization leads to two distinct minima (see e.g. Ref. 5 for more details). We can organize all local minima in a network by drawing a link between two minima if there is a saddle point with Morse index 1 between them. Such networks capture in a drawing the essentials of the topology of the landscape even when the dimensionality is high and we are not able to draw plots such as Fig. 1a any more.

Saddle points that lead after optimization to all minima of the toy model can be found by using a method called saddlepoint construction, described in Ref. 5 in the case of lens design and generalized in the Appendix in order to facilitate its use for the toy model⁸.

Note that for N=6 the Morse index 1 saddle point z = (1/4, 1/4, 1/4, 1/4, 1/4, -1/4) can be obtained by adding to the main hub for N=4, z = (1/4, 1/4, 1/4, 1/4), a new pair of coordinates $z_5 = 1/4$, $z_6 = -1/4$. The values of *f* given by Eq.(1) for the 6-dimensional saddle point and the 4-dimensional main hub are the same. Since the new pair of coordinates does not affect the value of *f*, it is called a null element. It can be easily seen that all fundamental Morse index 1 saddle points for N>2 can be obtained in the same way from the main hub with N-2 variables.

Since the toy model satisfies the mathematical requirements described in the Appendix, not only the main hub, but also the other four local minima for N=4 can be used to generate saddle points with N=6 by using saddle-point construction. By adding to the minimum z = (-Y, y, y, y), with Y=0.9205 and y=0.640167, a null element $z_5 = y$, $z_6 = -y$ we obtain a Morse index 1 saddle point z = (-Y, y, y, y, y, -y). By permuting the coordinates, we find 30 saddle points of this type. Other 60 saddle points that are permutations of z = (-Y, Y, -Y, y, y, y) (i.e. the null element is e.g. $z_1 = -Y$, $z_2 = Y$) will be neglected in what follows. They lead on one side of the saddle to the degenerate critical points mentioned above.

The network for the toy model is shown in Fig. 3. The 1+6+15=22 points denoted by "M" on the first, third and fifth row of the network are local minima and the 6+30=36 "S" points on the second and fourth row are saddle points. The two downward paths of local optimization started on both sides of a saddle point S_{i-j} lead to minima M_i and M_j, as indicated by the continuous lines between systems. The six saddle points on the second row are the fundamental critical points with Morse Index 1 that are permutations of z = (-1/4, 1/4, 1/4, 1/4, 1/4, 1/4). As in Fig. 1a, on one side they are all linked to the main hub M₁, here with z = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6), on the other side they are linked to their own projections on the third row M₂-M₇ that are permutations of z = (-Y, y, y, y, y, y) with Y=0.519464 and y=0.303893. The 30 saddle points that are permutations of z = (-Y, y, y, y, y, y) with Y=0.9205 and y= 0.640167 are shown on the fourth row and the minima M₈-M₂₂ that are projections of Morse index 2 critical points and are permutations of z = (-Y, -Y, y, y, y, y) with Y=0.847311 and y=0.673656 are on the last row.

The network in Fig. 3 is a perfectly regular arrangement. The regular pattern that can be easily observed in the first four rows includes in fact the fifth row as well, but in a two-dimensional drawing such as Fig. 3 this is less obvious. In fact, any third-row minimum has a link drawn upwards to the main hub and five drawn downwards that finally arrive at all other five third-row minima, with a minimum in the fifth row as an intermediate stage. The total number of 15 minima in the fifth row is exactly the number necessary to link any third-row minimum with all other five in the same row.

Note that from the five minima with N=4 (doublets), all local minima for N=6 can also be obtained in a way that is different from the one shown in Sec.4, by first constructing the saddle points with N=6 and then by optimizing these saddle points on both sides of the saddle. Similarly, the five minima with N=4 can be also obtained with saddle-point construction from the N=2 (singlet) main hub.



 $\overline{M8}$ $\overline{M9}$ $\overline{M10}$ $\overline{M11}$ $\overline{M12}$ $\overline{M13}$ $\overline{M14}$ $\overline{M15}$ $\overline{M16}$ $\overline{M17}$ $\overline{M18}$ $\overline{M19}$ $\overline{M20}$ $\overline{M21}$ $\overline{M22}$ Fig.3. Network of the toy model. Large points are local minima, small points are saddle points with Morse Index 1. The fundamental network for three lenses, in which the systems are obtained with ray tracing, has exactly the same structure. In the toy model, each minimum shown here is surrounded by six saddle points, but, as explained in the text, for each of the 15 minima in the fifth row four saddle points of the type z = (-Y, Y, -Y, y, y, y) linked to it, are omitted.



	Numerical MI 1	Toy Model Morse Index 1			Numerical Minimum	Toy Model Minimum	
S1-2				M2			
S1-3				М3			
S1-4				M4			
S1-5				M5			
S1-6				M6			
S1-7				M7			
a				b			

Fig.5. Second row of saddle points (a) and third row of minima (b)

	Numerical MI 1	Toy Model Morse Index 1		Numerical MI 1	Toy Model Morse Index 1	
S2-14			\$5-16			
S2-8			\$5-17			
S2-9			\$5-21			
S2-10			\$5-10			
S2-11			S5-18			
S3-19			\$6-13			
S3-8			S6-17			a
S3-15			\$6-15			
S3-12			\$6-22		-((-((-1)	
S3-18			S6-11			
S4-13			\$7-19			
S4-20			\$7-20			
S4-9			\$7-21			
S4-12			\$7-22			
S4-16			\$7-14			

Fig.6. Fourth row of saddle points

	Numerical Minimum	Toy Mo	odel Minimum	Toy Model Morse Index 2		
M8						
M9						
M10						
M11						
M12						
M13						
M14		$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left(\sum_{j$				
M15						
M16		-{}-{}-{}-{				
M17						
M18						
M19						
M20						
M21						
M22		-((-((-1)			40 	

Fig.7. Fundamental critical points with Morse index 2 and their projected minima on the fifth row. M15 is the Cooke Triplet shape. After reoptimization with variable distances between surfaces and adequate glasses, it leads to the well-known Cooke Triplet.

6. COMPARISON WITH THE FUNDAMENTAL TRIPLET NETWORK

As mentioned earlier, one of the goals of the present work is to find the simplest model that can explain the fundamental network for three lenses, where the variables of the global optimization problem are the surface curvatures¹. It turns out that the fundamental triplet network shown in Figure 3 of Ref. 1 is exactly the network shown in Fig. 3. Therefore, in the triplet design space, in the absence of the mechanism that is responsible for the destruction of local minima (e.g. when the distances between surfaces increase, see Fig. 5 of Ref. 1), the toy model captures the essentials of the topology of the merit function landscape based on ray tracing, and explains why we have exactly 22 system shapes there.

For comparison with numerical results¹, the z values of the critical points of the toy model must be converted to lens curvatures. This can be done in two different ways. Both of them produce results that are in good qualitative agreement with the numerical results given below, but each one has its own advantages and disadvantages at the level of quantitative details. A simple way is to identify z_k with the power of surface k. Since the origin of the toy model is a thin-lens (aberration) model, t in Eq.(2) becomes the total power of the system. The curvature of a surface is then given by

$$c_k = z_k / (n' - n) \tag{19}$$

where n is the refractive index after refraction and n is the refractive index before refraction.

In Figures 4, 5, 6 and 7 we use a different method, described in detail in Ref. 6, where it has already been used to associate curvatures to the fundamental critical points given by Eqs. (11) and (12). As shown there, the z values can be expressed in terms of paraxial ray angle changes at a surface, which can then be expressed in terms of surface curvatures. For simplicity, it is assumed that the imaging has transverse magnification -1 and that the three lenses have the refractive index 1.5.

Figures 4, 5, 6 and 7 show the systems in the fundamental triplet network¹, together with the corresponding system predicted by the toy model. In each case, the set of *z* values is shown as a bar chart. The toy-model system drawings are produced by substituting in the lens file for the numerical system the set of curvatures resulting from the *z* values in the corresponding bar chart. (Some artifacts, such as the excessive size of some lenses in these drawings, result from an insufficient control of the lens drawing process for the toy-model systems.) In Fig. 7, in addition to the local minima in the 5th-row of the network in Fig.3, the fundamental critical points with Morse Index 2 that produce the 5th-row minima via critical-point projection are shown.

We observe that the toy model gives a reasonable rough estimation of the system shape for the systems appearing in the fundamental triplet network. The lenses computed with the model have the correct type: positive, negative or menisci with positive or negative curvatures. Not surprisingly, there are significant quantitative differences between the model and numerical results. For instance, in many cases there is even a scale difference, i.e. for a better agreement the curvatures of the toy-model systems could be multiplied with a different, system-dependent factor. But considering the gross approximations that have been made to derive the toy model from aberration theory, a better quantitative agreement between the results of the model and the numerical ones cannot be expected. However, the triplet problem discussed here leads to a more general insight: the good qualitative agreement with the numerical systems obtained with lens design software shows that the toy model describes correctly the essence of the process of appearance of new local minima in the lens design landscape when the number of lenses in the design increases.

7. CONCLUSIONS

The growing diversity in the design landscape when the number of lenses increases can be understood in two different ways. The number of fundamental critical points increases rapidly with the dimensionality of the problem, and because of the one-to-one correspondence mentioned in Sec. 4, so does the number of system shapes. Alternatively, the number of local minima increases rapidly with the dimensionality of the problem, because there are so many ways to construct saddle points with M(=N/2) lenses from local minima with M-1 lenses. Saddle-point construction has already been used as a design tool in practical lens design problems^{5,9,10}. Rather than being just another addition to the long list of known empirical tools in lens design, saddle-point construction has a fundamental underlying basis, as shown by the toy model.

To the author's knowledge, models similar to the present one that explain how the complexity of a global optimization landscape increases with the dimensionality of the problem are not found in the literature. Lens design is a "real-world" problem described by the toy model, but, because it is mathematically so general, this model might be useful in other design or research areas as well.

8. APPENDIX: SADDLE-POINT CONSTRUCTION

Consider a smooth function $f(z_1, z_2, ..., z_N)$ with N variables, subject to the constraint $\sum_{i=1}^{N} z_i = t = const$. Assume that $\overline{M} = (\overline{z_1}, \overline{z_2}, ..., \overline{z_N})$ is a local minimum of f. We add now two new variables z_{N+1}, z_{N+2} and define a function $\hat{f}(z_1, z_2, ..., z_{N-1}, z_N, z_{N+1}, z_{N+2})$ subject to the constraint $\sum_{i=1}^{N+2} z_i = t = const$. Here, for simplicity the two new variables have been added at the end of the list of variables, but their position in the list can in fact be arbitrary. Eq. (1) is an example of a function f that can be transformed into a function \hat{f} of the same kind simply by changing N into N+2.

We will show below that from the local minimum \overline{M} of f in the space with N variables we can "construct" a saddle point with N+2 variables of \hat{f} if \hat{f} is such that we can use one of the new variables (e.g. z_{N+1}) in order to restore the "old" value $f(\overline{M})$ after any change of \hat{f} produced by two other variables (an old and the other new one).

Mathematically, saddle-point construction is stated as follows: Assume that there exist two transformations that leave \hat{f} invariant

1. $\hat{f}(\overline{z_1}, \overline{z_2}, ..., \overline{z_{N-1}}, \overline{z_N}, -p, p) = f(\overline{M})$ for any p and 2. $\hat{f}(\overline{z_1}, \overline{z_2}, ..., \overline{z_{N-1}}, q, -q, \overline{z_N}) = f(\overline{M})$ for any q.

Then, the point $\hat{f}(\overline{z_1}, \overline{z_2}, ..., \overline{z_{N-1}}, \overline{z_N}, -\overline{z_N}, \overline{z_N})$ is a critical point of \hat{f} . If it is non-degenerate, it is a saddle point with Morse Index 1.

The proof is shown below. A transformation of the type $z_i = z_i(p)$, i = 1...N + 2 with variable p defines a curve in the variable space. In the particular case of the two transformations mentioned above we have two straight lines. Since $z_1, z_2, ..., z_{N-1}$ remain constant in both cases, we focus on the three-dimensional variable subspace (z_N, z_{N+1}, z_{N+2}) . The two lines are then given by

1.
$$z_N = \overline{z_N} = const$$
, $z_{N+1} = -p$, $z_{N+2} = p$ and
2. $z_N = q$, $z_{N+1} = -q$, $z_{N+2} = \overline{z_N} = const$

2. $z_N = q$, $z_{N+1} = -q$, $z_{N+2} = z_N = const$ Note that for $p = q = \overline{z_N}$ these two lines intersect at a point S having

$$z_N = z_N, z_{N+1} = -z_N, z_{N+2} = z_N.$$

The two intersecting lines define a plane in the three-dimensional variable subspace (z_N, z_{N+1}, z_{N+2}) . Considering also

the old variables $z_1, z_2, ..., z_{N-1}$ kept here constant, this plane turns out to be the constraint plane $\sum_{i=1}^{N+2} z_i = t = const$.

This can be seen by shifting the coordinate system to S, by defining two vectors along lines 1 and 2, (0,-1,1) and (1,-1,0) respectively, and by computing their vector product, which turns out to be (1,1,1), the normal vector to the constraint plane in the three-dimensional variable subspace.

If in a plane, two lines, along which a function is constant, cross, then the crossing point is a critical point of this function (see e.g. Fig. 1a). In the constraint plane, S is a critical point of \hat{f} because the projection of the gradient on this plane, which cannot point in two different directions (i.e. those of the two normals to the lines 1 and 2), must be zero. In addition, because $z_1, z_2, ..., z_{N-1}$ are kept unchanged as for the minimum \overline{M} of f, the derivatives of \hat{f} with respect to them remain zero as well. We assume here that S is a non-degenerate critical point. (For a discussion of the degenerate case see Ref. 5 and Fig. 10 there.) In a two-dimensional problem, non-degenerate critical points are maxima, minima or saddle points. For maxima or minima, close to a critical point the contours along which a function is constant (the equimagnitude contours) are ellipses which reduce to a point at the critical point. For saddle points, the equimagnitude contours are two crossing lines. In the two-dimensional constraint plane of the three-dimensional variable subspace, the critical point S is a two-dimensional saddle point, for which the two crossing equimagnitude lines are line 1 and 2. In the two-dimensional constraint plane, this saddle point is a maximum in one direction (the downward direction), and a minimum in the perpendicular direction. (The third direction in the three-dimensional variable subspace (z_N, z_{N+1}, z_{N+2}) is blocked by the constraint and does not play a role in the determination of the Morse Index.) Since \overline{M} is a minimum in the "old" variable space $z_1, z_2, ..., z_{N-1}$ (we can consider them as independent variables, and assume that z_N is solved as in Eq.(6) in \overline{M} to satisfy the constraint) S is a non-degenerate critical point with only one downward direction, i.e. a saddle point with a Morse Index of 1.

Since the pair $z_{N+1} = -\overline{z_N}$, $z_{N+2} = \overline{z_N}$ that transforms the minimum \overline{M} into a saddle point does not affect $f(\overline{M})$, we call this pair of variables a "null element".

ACKNOWLEDGMENTS

The author thanks Pascal van Grol for valuable observations and help with Figures 3, 4, 5, 6 and 7 and Thomas Liebig for numerical runs of the toy model.

REFERENCES

[1] Van Grol, P., Bociort, F. and Van Turnhout, M., "Finding order in the design landscape of simple optical systems", Proc. SPIE **7428**, 742808, (2009).

[2] Bociort, F., Van Driel, E. and Serebriakov, A., "Networks of local minima in optical system optimization," Optics Letters **29**, 189-191 (2004).

[3] Gross H., Zuegge, H., Peschka, M., and Blechinger, F. [Handbook of Optical Systems Vol. 3] Wiley-VCH, 362-365, (2007).

[4] Marinescu, O. and Bociort, F., "Network search method in the design of EUV lithographic objectives," Applied Optics 46, 8385-8393 (2007).

[5] Bociort, F., and Van Turnhout, M., "Finding new local minima in lens design landscapes by constructing saddle points", Optical Engineering **48**, 063001 (2009).

[6] Bociort, F. and Van Grol, P., "Systematics of the design shapes in the optical merit function landscape," Proc. SPIE 7717, 77170 (2010).

[7] Gluss, D. and Weisstein, E., "Lagrange Multiplier", <u>http://mathworld.wolfram.com/LagrangeMultiplier.html</u>. [8] Additional information on saddle-point construction can be found on the web page

<u>http://wwwoptica.tn.tudelft.nl/users/bociort/networks.html</u>. The present address of the web page may change in the future, but using a web search with the page name "Networks, local minima and saddle points in optical system optimization" will lead hopefully to the new address.

[9] Marinescu, O. and Bociort, F., "Saddle-point construction in the design of lithographic objectives, part 1: method," Optical Engineering 47, 093002 (2008).

[10] Marinescu, O. and Bociort, F., "Saddle-point construction in the design of lithographic objectives, part 2: application," Optical Engineering **47**, 093003 (2008).