**Systematics of the design shapes in the optical merit function landscape**

F. Bociort* and P. van Grol
Delft University of Technology, Optics Research Group
Lorentzweg 1, NL - 2628 CJ Delft, The Netherlands

**ABSTRACT**

In this paper we describe new properties of the design landscape that could lead in the future to a new way to determine good starting points for subsequent local optimization. While in optimization the focus is usually only on local minima, here we show that points selected in the vicinity of other types of critical points (i.e. points where the merit function gradient vanishes) can be very useful starting points. We study here a problem that is simple enough to be analyzed in detail, the design landscape of triplets with variable curvatures. We show here how representatives of all triplet design shapes observed in global optimization runs can be obtained in a simple and systematic way by locally optimizing for each design shape one starting point obtained with the new method. Good approximations of these special starting points are also computed analytically with two theoretical models. We have found a one-to-one correspondence between the possible triplet design shapes and the critical points resulting from a model based on third-order spherical aberration within the framework of thin-lens theory. The same number and properties of critical points are predicted by a second model, which is even simpler and mathematically more general.

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

1. **INTRODUCTION**

Typical design landscapes for imaging systems have multiple local minima, and when the number of design parameters grows the number of local minima increases rapidly. Knowing what types of local minima can be expected can be very useful for global optimization. In this paper we examine the design landscape of triplets with curvatures as variables, for which the design landscape is already fairly complex, but still manageable. An earlier study indicates that such a merit function landscape consists of main valleys, separated by higher merit function barriers, e.g. the valleys A and B in Fig. 1. Within the main valleys, called “design shapes”, occasionally several minima with comparable merit function, separated by low barriers, can be found, e.g. the minima A1 and A2 in valley A. Ignoring at this stage the possible presence of several similar designs in the same main valley, our goal will be to find in an effective way all possible system shapes in our triplet landscape.

Considering the facts that the merit function is a nonlinear function of several design parameters and that it is usually computed on the basis ray tracing, it is not surprising that finding many different local minima in lens design cannot be achieved with analytic methods. However, we will show that at, least in the triplet landscape, the most difficult part of this problem, finding adequate starting points for subsequent local optimization, can be done analytically. From a set of starting points computed analytically, after local optimization with a usual numerical merit function all possible triplet shapes, including the well-known Cooke triplet, will be obtained. For triplets, good starting points for designing e.g. the Cooke triplet can be obtained by using aberration theory but our main purpose here will be to obtain insight that will lead hopefully to new design methods for more complex systems.

In Section 2 we will discuss the concept of critical point and in Sec 3 we will present results of numerical experiments showing how all possible triplet shapes can be obtained from points selected in the vicinity of critical points. In Sec. 4,
these special critical points will be calculated analytically, and in Sec. 5 we will see how all tripled design shapes can be obtained by locally optimizing starting points derived from the amazingly simple analytical results of Sec. 4.

Fig. 1. Two main valleys, A and B, called system shapes. Two local minima, A1 and A2, with a low merit-function barrier between them, form valley A.

2. CRITICAL POINTS

In our task to find the possible triplet system shapes (and also, we would argue, in analyzing global optimization problems in general) it is useful to think in terms of basins of attraction (See Fig. 2a). The set of all starting configurations that lead to the same minimum after local optimization is called the basin of attraction for that minimum. However, since any basin point can be used for obtaining a given minimum, in the present work we will choose the basin points that can be determined in the easiest way. We will see in the next section that special points on the basin borders can be used for producing good starting points. These special points on the borders are saddle points of the merit function (the large dots in Fig.2 a,b). For triplets, as shown in Sec. 4, it is possible to compute analytically sufficiently accurate values for their set of parameters.

Fig.2 Basins of attraction. a) For obtaining a certain local minimum it is sufficient to choose a starting point in its basin of attraction and to locally optimize it. Saddle points are critical points on basin borders (the large dots). Points chosen in their immediate vicinity lead to a) two or b) more local minima after subsequent local optimization.

For simplicity, we first consider a global optimization problem with no constraints. (A constraint case will be discussed later in this paper.) A point in the design space is described by the vector \( \mathbf{x} = (x_1, x_2, \ldots, x_N) \) whose components are the \( N \) optimization variables. The critical points in the \( N \)-dimensional design space are the points for which the gradient of the merit function \( f \) vanishes

\[
\nabla f(\mathbf{x}) = \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_N} \right) = 0
\]

(1)

In a small neighborhood around a critical point with merit function \( f_0 \), a quadratic approximation for \( f \) can be used

\[
f(\hat{\mathbf{x}}) = f_0 + \sum A_{ij} \hat{x}_i \hat{x}_j
\]

(2)

where the circumflex denotes the optimization variables in a translated coordinate system that has its origin at the critical point, and where \( A_{ij} \) are the elements of the matrix of the second-order derivatives with respect to the optimization
variables, computed at the critical point. As known from linear algebra, the coordinate system can be rotated in such a way that the quadratic form in Eq.(2) contains only squares of the variables (denoted below by a bar) in the new coordinate system. Eq.(2) now becomes

\[ f(\bar{x}) = f_0 + \sum \lambda_i \bar{x}_i^2 \]  

(3)

The axes of the new coordinate system are then oriented along the eigenvectors of matrix \( A \), (i.e. \( \bar{x}_i \) are measured along these eigenvectors) and the factors \( \lambda_i \) in Eq.(3) are the corresponding eigenvalues. Assuming that all eigenvalues are nonzero, the Morse index (MI) of the critical point is defined as the number of negative eigenvalues in Eq. (3)\(^3\). As can be seen by putting in Eq.(3) \( \bar{x}_i = 0 \) for all \( i \neq j \), a negative eigenvalue \( \lambda_j \) means that along the direction defined by the corresponding eigenvector (i.e. by varying \( \bar{x}_j \)) the critical point is a maximum, and a positive \( \lambda_j \) indicates a minimum along direction \( j \). Local minima have MI=0 (all eigenvalues are positive), i.e. they are minima in all directions, and local maxima have MI=N. For saddle points, the Morse index has values between 1 and \( N-1 \). In the analytic and numerical results discussed in this paper we will encounter critical points with Morse indices 0, 1 and 2.

Fig. 3. Saddle point with Morse index 1. The direction of the eigenvector having the negative eigenvalue is shown in green. Only one of the N-1 directions with positive eigenvalues is shown here in red.

Figure 3 shows a saddle point with MI 1. Points chosen in its immediate vicinity lead after subsequent local optimization to two local minima on opposite sides of the saddle. The large dot in Fig. 2a shows the two corresponding basins of attraction. In the immediate neighborhood of a MI 2 saddle point we can find a plane (defined by the two directions with \( \lambda_i<0 \)) in which the saddle point is a maximum. As can be understood intuitively by thinking about a hill top surrounded by several valleys in a two-dimensional landscape, in general more than two local minima can result by locally optimizing points close to a MI 2 saddle point (the large dot in Fig. 2b).

3. TRIPLET DESIGN SHAPES

The systems \( M_i \) in Fig. 4 are the design shapes in a global optimization search which has been discussed in detail in Ref. 1. (The numbering of the minima in Fig. 4 is the same as in Fig. 3 there.) The image defects in the merit function are transverse ray aberrations (computed with respect to the chief ray) and our merit function in the two optimization examples in this paper is the default error function of the software we have used, CODE V. For simplicity, edge thickness control has been disabled in these runs. Each system has three lenses with the same glass SK16 (Schott) and glass thickness of 1mm, placed at two equal distances of 1.5mm from each other. The transverse magnification is -1, the image is at the paraxial position, there is no vignetting, the entrance pupil diameter is 9mm, there are three fields (0, 7.14 and 10 degrees) and three wavelengths (corresponding to the standard F, d, and C visible spectral lines, which are in fact not necessary because chromatic aberration correction is not possible with only one glass type, but which are used in order to have specifications that do not differ unnecessarily from those used in the next example, which is a realistic one). The six surface curvatures are variable. Since there is a constraint on total track (equal to 120 mm), the search space is effectively five-dimensional. This set of specifications has a special property that will be apparent later.
Fig. 4 The 22 fundamental system shapes $M_i$ for triplets. All system shapes can be obtained with local optimization from the MI 1 critical points $C_2 – C_7$ and from the MI 2 critical points $C_8 – C_{22}$. 
There are 29 local minima in this run, which correspond to 22 system shapes. All these system shapes have been found with our own global optimization program NETMIN\textsuperscript{4,5,6}, 21 of them could also be obtained with CODE V’s Global Synthesis. We have found no additional design shapes with Global Synthesis. As mentioned in Ref. 1, different triplet global optimization runs have been performed by changing aperture, field, transverse magnification, distances between surfaces, glasses etc, but the distances between lens surfaces were in all cases not too large. We call the set of design shapes shown in Fig. 4 “fundamental” in the sense that any system shape observed in such a triplet run was one of the shapes of this fundamental set, and no other system shapes could be ever observed. In the next chapter the number of 22 system shapes will follow from a theoretical model discussed there. M\textsubscript{15} is the well-known Cooke triplet shape. The set of specifications of this run has been determined by several trials in such a way that all system shapes can be found in a single global optimization run.

Figure 4 also shows six MI 1 critical points C\textsubscript{2} –C\textsubscript{7} and fifteen MI 2 critical points C\textsubscript{8} –C\textsubscript{22}. Note the resemblance between the local minima M\textsubscript{i} and the corresponding saddle points C\textsubscript{i}. Both sets of systems contain meniscus lenses or meniscus-shaped airspaces, and the main difference between the two types of systems is that the meniscus curvatures are larger in the M\textsubscript{i} systems than in the corresponding C\textsubscript{i}. Excepting M\textsubscript{1}, we have a one-to-one correspondence between the M\textsubscript{i} systems and the corresponding C\textsubscript{i} and it is therefore not surprising that the 21 system shapes M\textsubscript{i} can be obtained via local optimization from the corresponding C\textsubscript{i}. For reasons that will become apparent in the next section, M\textsubscript{1} can be considered as a critical point C\textsubscript{1}.

The six MI 1 saddle points C\textsubscript{2} –C\textsubscript{7} lead on one side of the saddle to the corresponding M\textsubscript{i} (e.g. Lm\textsubscript{2} in Fig. 2a). On the other side, all these six MI 1 saddle points lead to the same minimum, M\textsubscript{1} (e.g. Lm\textsubscript{1} in Fig. 2a). A typical behavior for the fifteen MI 2 saddle points C\textsubscript{8} –C\textsubscript{22} is shown in Fig. 5 for one of them, C\textsubscript{10}. Depending on how points close to it are chosen, these points can lead after local optimization to four distinct system shapes (Lm\textsubscript{1}, Lm\textsubscript{2}, Lm\textsubscript{3}, and Lm\textsubscript{4} in Fig. 2b) One of them is the corresponding M\textsubscript{10} and the other three are system shapes that have already be obtained from the MI 1 saddle points. The one-to-one correspondence should be therefore understood in the sense that every C\textsubscript{i} produces a new M\textsubscript{i} which was not previously obtained via the same correspondence from critical points having a lower MI. Other critical points also exist in the triplet landscape\textsuperscript{1}, but because of this one-to one correspondence with the fundamental design shapes M\textsubscript{i}, C\textsubscript{i} will be called fundamental critical points.

Saddle points have a zero merit-function gradient and indeterminate gradient direction, and are therefore not adequate starting point for local optimization by themselves. However, adequate starting points in their neighborhood can be derived from them. A simple strategy that is successful in most cases is the following: in order to obtain a starting point that leads after optimization to the corresponding M\textsubscript{i} one should simply increase in C\textsubscript{i} the curvatures of the menisci. The menisci of such a starting point are shown dashed in Fig. 5. However, some precautions must be taken in local optimization, as shown in Sec 5. For the MI 2 saddle points, two-step strategies, with MI 1 saddle points as intermediate results, are also possible.

The fundamental critical points C\textsubscript{i} have been obtained from local minima of simpler problems having the same aperture and field specifications by using our so-called “saddle-point construction” (SPC) method\textsuperscript{1,7,8,9}. For more details on the general version of SPC we refer to these papers. The MI 1 critical points C\textsubscript{2} –C\textsubscript{7} have been obtained with SPC from a doublet minimum (m1 in Fig. 1 of Ref. 1) by inserting a meniscus at different positions (see s1-2…s1-7 in Fig. 3 of Ref. 1). Using essentially the same method\textsuperscript{1}, critical points with a higher MI can also be generated by inserting several meniscus lenses and meniscus airspaces in existing local minima. The MI 2 critical points C\textsubscript{8} –C\textsubscript{22} are obtained with SPC from a singlet, which strongly resembles the bi-convex lens in e.g. C\textsubscript{8}, C\textsubscript{9} and C\textsubscript{12}.

As a technical detail, the curvatures shown in Fig. 4 for the MI 1 and MI 2 critical points C\textsubscript{i} are in fact computed for lenses in which one or more distances between surfaces are zero, as required by SPC. The critical points continue to exist when these zero distances are increased to the values shown in Fig. 4, but their curvatures slightly change. The MI 1 critical points C\textsubscript{2} –C\textsubscript{7} can also be calculated with our program NETMIN. While the NETMIN computation is more time consuming, it has the advantage that it is directly applicable for lenses with distances between surfaces as shown in Fig. 4. A comparison for C\textsubscript{2} –C\textsubscript{7} between the approximate SPC results and the more accurate NETMIN results shows that the SPC results provide a very good approximation of the actual critical point curvatures, as shown in Fig. 6 for C\textsubscript{2}. We expect that for the MI 2 critical points C\textsubscript{8} –C\textsubscript{22} that exist in the landscape, the SPC results shown in Fig. 4 also provide a good approximation.
Fig. 5 Obtaining from a MI 2 critical point $C_i$ ($i=8\ldots22$) the corresponding system shape $M_i$. Here, $i=10$. The system with dashed lenses is the starting point for local optimization. The two dashed lenses are the menisci for which the curvatures have been increased. In addition to $M_{10}$, other system shapes are also obtained.

Fig. 6. $C_2$ with SPC (upper system) and NETMIN (lower system)

Rather than going into the details of NETMIN or SPC computations, what must be emphasized here is the fact that the fundamental critical points exist is the same triplet landscape (determined by ray tracing) as the triplet local minima. While Fig. 4 is a result of numerical experiments, and the theoretical reasons for the one-to-one correspondence mentioned above still need to be investigated, the idea presented here is attractive because all system shapes have been obtained in a systematic way from critical points that can be obtained easily. In the next section we will show how a very approximation for the fundamental critical points of the triplet landscape can be calculated analytically.

4. ANALYTIC MODELS FOR THE FUNDAMENTAL CRITICAL POINTS

While the numerical merit function of CODE V based on ray tracing is accurate and adequate for practical purposes, simplified models are better suited for obtaining insight into the properties of the fundamental critical points. We will use below two different simplified models. Remarkably, we will see below that the fundamental critical points will continue to exist even when very gross approximations are made. With these approximations, the models will become simple enough to be solved analytically.

An analytic approximation of the merit function can be obtained by writing a power series for the transverse aberration of each ray in the merit function in terms of the normalized aperture and field coordinates of the ray and of aberration coefficients. For our purposes even the gross approximation of keeping in the merit function only the contribution resulting from the third-order spherical aberration $SA$ will turn out to be sufficient. By writing

$$f = SA^2$$

the critical points satisfying the condition $\nabla f = 2SA\nabla SA = 0$ are solutions of either $\nabla SA = 0$ or of $SA = 0$. Consider a system with $L$ lenses and $N=2L$ spherical surfaces having variable curvatures $c_k$. We then have

$$SA = -\sum_{k=1}^{N} A_k^2 h_k \left( \frac{u_{k+1}}{n_{k+1}} - \frac{u_k}{n_k} \right)$$

where for our purposes it is convenient to write $A_k$ as

$$A_k = \frac{u_{k+1} - u_k}{n_{k+1} - n_k^{-1}}$$
Here $u_k$ is the angle of the paraxially traced marginal ray and $n_k$ is the refractive index, both before surface $k$. If the paraxial refraction formula

$$n_{k+1}u_{k+1} = n_ku_k - h_kc_k\left(n_{k+1} - n_k\right)$$

(7) is used to eliminate $u_{k+1}$ in Eq. (6), then the refraction invariant $A_k$ can be written after simple algebra in the well-known form $A_k = n_k\left(h_kc_k + u_k\right)$. However, in what follows we will use for $A_k$ the form given by Eq. (6), because then $SA$ can be expressed entirely in terms of angles $u_k$. We will use these angles as independent variables and when the critical points are found, their set of curvatures be determined from the angles with Eq. (7).

Consider a system of thin lenses in contact, in which all lenses are in air and have the same refractive index $n$. Therefore $n_k = 1$ for odd $k$ and $n_k = n$ for even $k$. For the height of the marginal ray we have $h_k = h = \text{const}$ for all $k$. Assume that the system has the given angles $u_1 = \alpha$ in object space and $u_{N+1} = \beta$ in image space. It will turn out that the fundamental critical points are the solutions of $\nabla SA = 0$ with $u_1, u_2, \ldots, u_N$ as variables. We have then N-1 equations $\partial SA / \partial u_k = 0$. Note that for each partial derivative of $SA$, only two consecutive terms from Eq (5) give nonzero contributions.

$$\frac{\partial}{\partial u_k} \left[ \left( \frac{u_k - u_{k-1}}{n_k - n_{k-1}} \right) \left( \frac{u_k - u_{k-1}}{n_k - n_{k-1}} \right) + \left( \frac{u_{k+1} - u_k}{n_{k+1} - n_k} \right) \left( \frac{u_{k+1} - u_k}{n_{k+1} - n_k} \right) \right] = 0$$

(8)

After a few calculations it turns out that the partial derivatives of these cubic terms are quadratic expressions that can be written as a product of two terms that are linear expressions in the variables $u_k$. Eqs (8) can then be written as

$$a_k b_k = 0$$

(9)

where

$$a_k = u_{k-1} - u_{k+1}$$

(10)

and

$$b_k = p_k\left(u_{k-1} + u_{k+1}\right) - 2q_ku_k$$

(11)

In the case of the present spherical-aberration model we have $p_k = n + 2$; $q_k = 2n + 1$ if $k$ is odd, and $p_k = 2n + 1$; $q_k = n + 2$ if $k$ is even.

As we will see below, it is useful to consider the more general class of problems, which is defined by the property that equations (9-11) are valid for the members of the class. The spherical-aberration model (4-6) is then a special case in this class. The simplest member of this class is a purely cubic model which will be called in what follows the “toy” model:

$$f = T^2$$

(12)

with

$$T = \sum_{k=1}^{N} (u_{k+1} - u_k)^3$$

(13)

In this case, the equations $\partial T / \partial u_k = 0$ lead to equations (9-11) with $p_k = q_k = 1$ for all $k$. Note that the refractive index does not appear in the toy “merit function”. While from the point of view of aberration theory Eqs. (12-13) may therefore seem irrelevant, it turns out that this gross simplification can still produce a useful approximation for the variables $u_k$ of the critical points. We will see that the toy model (12-13) has the same topology as the thin-lens spherical-aberration model and that the toy model captures the essentials of the topology of the merit function landscape based on ray tracing in the design spaces we consider here.
To solve a system of equations of the type (9), note that for each \( k=2,3\ldots N \) the corresponding equation is satisfied either if \( a_k = 0 \) or if \( b_k = 0 \). Thus, we can construct \( 2^{N-1} \) systems of linear equations: In each such system, there is one equation for each \( k \), either \( a_k = 0 \) or \( b_k = 0 \). Each critical point will be a solution of one of the systems of linear equations. However, some of these linear systems of equations will turn out to be incompatible; therefore the number of critical points will be less than \( 2^{N-1} \).

In the triplet case we have \( L=3 \) and \( N=2L=6 \) and the unconstrained critical-point problem with given angles \( u_i = \alpha \) in object space and \( u_i = \beta \) in image space, and five variables \( u_2, u_3, u_4, u_5, u_6 \) leads, both for the spherical-aberration model and for the toy model, to 32 systems, each with five linear equations of the type \( a_k = 0 \) or \( b_k = 0 \). It turns out that 10 systems are incompatible and have therefore no solution. The remaining systems can be easily solved and the solutions are 22 critical points. For each solution, the Morse index has been determined by using the procedure described in Sec. 2. The five eigenvalues \( \lambda_i \) that appear in Eq.(3) are computed by standard linear algebra methods and turn out to be proportional with \( \alpha - \beta \). Assuming \( \alpha > \beta \) (for real imaging the marginal ray has negative angle in the image space and positive or zero angle in object space), the MI of each critical point is the number of negative eigenvalues in the corresponding set of five numbers. It turns out that we have one critical point with MI 0, 6 critical points with MI 1 and 15 critical points with MI 2.

Both for the spherical aberration model and for the toy model, the set of six lens curvatures for each critical point can be determined from the seven angles by using Eq. (7). Considering the case with transverse magnification \( TM= -1 \), by choosing \( \alpha = 1, \beta = -1 \) and taking for simplicity \( n=1.5=3/2, h=1 \) we obtain the curvatures given in Table 1 as integers or quotients of two integers.

As an example, for the MI 2 critical point with number 15 that corresponds to the Cooke Triplet design shape, the five equations of the spherical aberration model are

\[
4(1+u_3) - 7u_2 = 0, u_2 - u_4 = 0, 4(u_3 + u_5) - 7u_4 = 0, u_4 - u_6 = 0, 4(u_5 - 1) - 7u_6 = 0
\]  

(14)

The second and fourth equations in (14) result from Eq.(10), the other three equations result from Eq.(11) with \( k \) even. The solution is easily obtained as

\[
u_2 = u_4 = u_6 = 0, u_5 = -u_3 = 1 \]

(15)

The curvatures resulting from Eqs. (15) and (7) are given in line 15 of Tab.1. The angular variables and curvatures for the other critical points are obtained in the same simple way.

For the same special case and critical point with number 15, the five equations of the toy model are

\[
1 + u_3 - 2u_2 = 0, u_2 - u_4 = 0, u_3 + u_5 - 2u_4 = 0, u_4 - u_6 = 0, u_5 - 1 - 2u_6 = 0
\]  

(16)

While this is not the case for all critical points, for this critical point the solutions of the toy equations (16) are the same as for the spherical aberration equations (14) and are given by Eqs. (15).

For the toy model the angular variables can also be determined in an equivalent way that offers more insight into the properties of the solutions. By changing the variables to

\[
z_k = u_{k+1} - u_k
\]  

(17)

the unconstrained problem of finding the critical points of Eq.(13) with angular variables becomes the constrained problem of finding the critical points of

\[
T = \sum_{k=1}^{N} z_k^3
\]  

(18)

subject to the constraint

\[
\sum_{k=1}^{N} z_k = \beta - \alpha = t
\]  

(19)
To transform this constrained problem into an unconstrained one we use the well-known Lagrange multiplier method \(^{12}\). With \(\mu\) as the so-called Lagrange multiplier, we now look for the critical points of the Lagrange function associated to \(T\)

\[
T^* = \sum_{k=1}^{N} z_k^3 - \mu \sum_{k=1}^{N} z_k
\]  

(20)

From the equations

\[
\frac{\partial T^*}{\partial z_k} = 3 z_k^2 - \mu = 0
\]  

(21)

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

\[
|z_k| = \sqrt{\frac{\mu}{3}} = \omega
\]  

(22)

and we can write

\[
z_k = \pm \omega
\]  

(23)

The set of critical points is obtained by permuting the plus and minus signs in all possible ways, and \(\omega\) is chosen so that the constraint (19) is satisfied. The critical points can then be classified according to the number negative signs in the solution vector \(z = (z_1, z_2, ..., z_N)\). By changing the solutions back to angular variables the Morse Index can be computed as shown above, and it turns out that the number of minus signs in the solution vector of a given critical point is exactly its Morse index. We therefore have

\[
\omega_{MI} = t / \left( N - 2MI \right)
\]  

(24)

All values of \(MI < N/2\) need to be considered.
Fig. 7 The 22 fundamental critical points $C_i$ computed with the toy model for triplets. For comparison, the overlap with the critical points $C_i$ of the numerical merit function shown in Fig. 4 (here in red, dashed) is also given. The overlap is so good that the two drawings can barely be distinguished.
For triplets, there is one solution with MI 0 given by the vector
\[ z = (t/6, t/6, t/6, t/6, t/6, t/6) \]  \hfill (25)
Depending on the position \( k \) where the minus sign is placed, there are 6 solutions with MI 1, e.g. for \( k = 1 \) we have
\[ z = (-t/4, t/4, t/4, t/4, t/4, t/4) \]  \hfill (26)
By permuting two minus signs over six positions, we find \( 6^5 \times 2 = 15 \) solutions with MI 2, e.g. the one corresponding to critical point 15
\[ z = (t/2, t/2, -t/2, -t/2, t/2, t/2) \]  \hfill (27)
Changing the solutions (27) back to angular variables with Eq.(17) leads to Eqs. (15), as expected. As can be easily seen, in all solutions of the type (25), (26) or (27) summing up the six vector components gives \( t \), as expected from Eq.(19).
The total number of critical points predicted by the toy model for the triplet problem is thus 22, the same as the number of fundamental critical points of the CODE V merit function computed with SPC.

For comparison with the 22 fundamental critical points \( C_i \) of the numerical merit function, the curvatures of the critical points of the toy model in Tab. 1 must be divided by a scale factor \( s \). For the example shown in Fig. 4 we have \( s = 66.4078 \). Table 2 gives the comparison for critical point 15 (that leads to the Cooke triplet shape) between the curvatures computed with SPC for the numerical merit function and the analytical curvatures. Note that, as shown in line 15 of Tab. 1 for this critical point the spherical aberration model and toy model produce the same values. For three curvatures there is a difference between SPC and the analytic values of about 20\%, for the other three the agreement is very good.

<table>
<thead>
<tr>
<th>( C_1 )</th>
<th>( C_2 )</th>
<th>( C_3 )</th>
<th>( C_4 )</th>
<th>( C_5 )</th>
<th>( C_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPC</td>
<td>0.029448</td>
<td>-0.025753</td>
<td>-0.025753</td>
<td>0.029448</td>
<td>0.0241551</td>
</tr>
<tr>
<td>analytic</td>
<td>0.0301169</td>
<td>-0.0301169</td>
<td>-0.0301169</td>
<td>0.0301169</td>
<td>0.0301169</td>
</tr>
</tbody>
</table>

Table 2. SPC and analytic values for the curvatures of \( C_{15} \) and \( C_{15}^* \)

Figure 7 shows the system drawings for the critical points of the toy problem and the comparison between the system drawings of the toy and SPC critical points. The overlap of system drawings cannot reveal adequately differences of the order of magnitude of those visible in Tab. 2, but it clearly indicates that the SPC and toy results correspond to precisely the same critical points. We conclude therefore that the results shown in Tab. 1 for the two models are in fact analytic approximations of the fundamental critical points shown in Fig. 4. For the toy model the fundamental critical points are denoted by \( C_i^* \). Note also that the two analytic models predict the existence of the MI 0 minimum \( C_1^* \) that corresponds to the CODE V minimum \( M_1 \). Therefore, this special minimum is considered in Fig. 4 as a critical point \( C_1 \). The overlap of system drawings between the SPC results and the analytical results of the spherical aberration model is similar to the overlap in Fig. 7 and is not shown here. In the above example we had \( TM = -1 \), but the SPC method and the two analytic models are applicable for any other value of transverse magnification as well. For instance, the same calculations for \( u_1 = \alpha = 0 \) give the analytical critical points of problems with object at infinity. The number of critical points does not depend on \( \alpha \) and \( \beta \).

Note that, with the exception of \( C_{15} \), in all other critical points in Tab. 1 we can find, with both analytic models, one or more meniscus lenses or meniscus-shaped airspaces. In most cases, the presence of a meniscus lens or airspace at a given position in the system is indicated by two consecutive curvatures that are equal. However, there are three more complex MI 2 cases. For instance, for \( C_{15}^* \) the lens meniscus with \( c_1 = c_4 = 2 \) is split by the air meniscus with \( c_2 = c_3 = 2 \). The other such cases are \( C_{14}^* \) and \( C_{16}^* \) where an air meniscus also splits a lens meniscus. However, \( C_{15}^* \) is special because its construction can be interpreted in two different ways; the other is that the lens meniscus with \( c_1 = c_6 = 2 \) is split by the air meniscus with \( c_3 = c_4 = 2 \).

With an early version of the present spherical aberration model, the presence of meniscus lenses has been studied for MI 1 saddle points\(^{13}\). The SPC method has been in fact developed by starting from this observation. This method has become a practical design tool to find new design shapes in a computationally efficient way and was successfully used in the design of state-of-the-art lithographic objectives\(^{6,15}\).

In this paper we discuss triplets. For the simpler problem of doublets, similar results have been discussed earlier. For doublets, there are five fundamental critical points in the numerical merit function landscape, four MI 1 saddle points
(the four saddle points shown e.g. in Fig. 1. of Ref. 1) and a MI 0 minimum (m1 in Fig. 1. of Ref. 1). For the case of an object at infinity, \( u_0 = \alpha = 0 \), the curvature values for the four MI 1 saddle points that result from the present spherical aberration model have already been computed in Ref. 13 and are listed in Table 1 there.

5. ANALYTIC STARTING POINTS FOR LOCAL OPTIMIZATION

In the same way as for the numerical SPC results, the analytic values for the fundamental critical points can be used to derive from them starting points for numerical optimization that lead to the 22 design shapes, according to the one-to-one correspondence shown in Fig. 4. However, for deriving these starting points it is important to identify the meniscus surfaces of the given critical point, as shown above. In the present problem with transverse magnification -1, a possible strategy is simply to increase the curvatures of all such meniscus surfaces, as mentioned earlier. If surface \( k \) is such a meniscus surface, then the curvature for the starting point is related to the analytical one by

\[
C_{k,\text{start}} = C_k + \epsilon
\]

Eq. (28) must be applied to all meniscus surfaces of the given critical point. In other cases, e.g. when the object is at infinity, a more elaborate strategy is needed.

Curvatures of other surfaces can be adapted so that constraints are satisfied. The value of \( \epsilon \) must not be too small, for two reasons. First, MI 1 and MI 2 saddle points are points on basin borders, as shown in Fig. 2. For good results, the starting points for local optimization should be deep enough within the basin, so that the merit function gradient for these points is sufficiently large for stable numerical behavior. Also, note that, because the analytic values are approximations, the resulting point is not exactly on the basin border of the numerical merit function. However, with an \( \epsilon \) that is large enough, the starting point is placed in the correct basin of attraction for the one-to-one correspondence.

An additional practical issue to be solved is that, for achieving the expected one-to-one correspondence systematically, the behavior of the local optimization algorithm as it is implemented at present in the software we use and in other commercially available programs must be changed. Most optical design programs use for local optimization elaborate versions of the Levenberg-Marquardt damped least squares method in which the damping factor is determined automatically. However, in order to increase computational efficiency, the automatically determined damping is often very low. As shown earlier, in this case neighboring basins of attraction can be mixed together and can form very complex patterns\(^{16,17}\). Also, while for the type of merit function used in optical design, Levenberg-Marquardt and related methods are an excellent choice for starting points close enough to minima, there are situations when this type of method is not optimal\(^{18}\). For starting points close to saddle points, a steepest descent method would be preferable.

Both problems are solved when damping is increased sufficiently. Since in the present version of CODE V the user cannot control the damping factor directly, we have used for the first cycles of local optimization a CODE V macro that increases the damping externally. In the final stages, standard CODE V optimization is used. Before each optimization cycle, the values of the optimization variables are read in the macro. These values are the components of a vector denoted by \( x_0 \). The macro uses the standard technique to add at each iteration a penalty term to the merit function that discourages large changes of variables

\[
f_D(x) = f(x) + d(x - x_0)^T(x - x_0)
\]

When the damping factor \( d \) is sufficiently large, the basin borders have a desirable regular shape like the one shown in Fig. 2 and the direction of optimization comes close to the direction of steepest descent.

Despite the gross approximations made in the derivation of the toy model, from each analytic critical point in Tab. 1 the corresponding design shape can be obtained automatically by performing the steps described in this section. We have written a second macro, which reads a lens file having the correct specifications in terms of aperture, field, wavelengths, distances between surfaces, glasses etc, then reads successively the sets of six curvatures for each analytical critical point and then performs first damped then standard optimization. With this macro that is available via our web site\(^{19}\) all 22 design shapes \( M_i \) in Fig. 4 are found without user intervention in a few minutes from starting points derived from the simple data given in Tab.1.
The specifications of the previous triplet global search problem given in Sec. 3 are interesting for theoretic, rather than practical, reasons: the 22 resulting system shapes match exactly the theoretical predictions of the two analytic models discussed in Sec. 4. However, the same approach for obtaining all system shapes from starting points computed analytically also works for a triplet global search with realistic Cooke triplet specifications.

The specifications of a Cooke triplet lens file that comes together with CODE V (object at infinity, f/4.5, field of 20 degrees, polychromatic imaging) have been used for a global search, for which the variables are the surface curvatures and the effective focal length is kept constant to 50 mm. Earlier numerical results with our program NETMIN and with Global Synthesis of CODE V show that only 15 of the 22 system shapes of the previous global search are found in the Cooke triplet global search. Seven system shapes, M1, M3, M6, M12, M16, M20, and M22 are absent, and no new system shapes are present1. Apart from the two larger airspaces, the 15 system shapes that are found resemble strongly the corresponding system shapes in Fig. 4. Note that the spherical aberration model described in Sec. 4 uses a thin-lens approximation, in which all glasses are equal. Since the distances between lenses are smaller and only one glass is present, the previous global search comes closer to the requirements of this analytic model than the Cooke triplet global search. It is therefore not surprising that the results of the Cooke triplet global search (i.e. the absence of seven system shapes) differ more from the predictions of the spherical aberration model than the results of the previous global search.

For an object at infinity, \( u_1 = \alpha = 0 \) and otherwise the same settings as in Sec.4, a set of 22 critical points has been generated with the toy model, and the results are as simple as those shown in Tab. 1 for the case TM=−1. For instance, the critical point \( C^*_{15} \) that corresponds to the Cooke triplet has the set of six curvatures \((1.5, -0.5, -0.5, 1.5, 1.5, -0.5)\). By loading in the same macro as in the previous case the Cooke triplet lens file that comes together with CODE V and the curvatures of the critical points from the toy model (divided by a scale factor \( s=75.88 \)), all design shapes of the Cooke triplet global search are obtained automatically. (In some cases a strategy more general than Eq. (28) is needed.) For the 15 critical points for which the corresponding systems shapes exist, our modified optimization procedure produces the expected result as in the previous example. From each of the seven critical points for which the corresponding system shapes does not exist, another system shape from the set of 15 existing ones is obtained. Figure 8 shows how a Cooke triplet design is obtained from the critical point \( C^*_{15} \) computed analytically.

![Figure 8](image)

6. CONCLUSIONS

Global optimization is essential in optical design and in other engineering areas. However, for practical applications the complexity of the problem is in most cases so daunting that designers give up hope to understand the properties of the corresponding design landscapes. The emphasis in this paper was on understanding the design landscape a problem that is simple enough for that purpose, rather than on practical goals.

Our goal was to find starting points for subsequent local optimization with a numerical merit function that is used for practical design, and to obtain from these starting points all types of local minima for the given specifications (by ignoring similarly looking solutions with nearly the same merit function). We have seen that there exist global
optimization problems with non-trivial topography (triplet problems with a number of minima of the order of 20) for which such starting points can be derived analytically.

In this case, there exist critical points of the numerical merit function that are in a one-to-one correspondence with the design shapes of the global optimization problem, and sufficiently accurate values for the set of variables of these points can be calculated with very simple analytic models (see e.g. Eqs. (25-27)). This property shows the presence of a high degree of order in the merit function landscape, and in future research it is worth investigating how far one can go in using this insight in order to develop practical tools for global optimization in complex design tasks.

ACKNOWLEDGMENTS

We thank Maarten van Turnhout for valuable discussions and for writing the CODE V macro for increasing damping during optimization. The use of a CODE V educational license is gratefully acknowledged.

REFERENCES

[19] The file systematics.zip containing CODE V macros and lens files is available from our web page. The present address http://wwwoptica.tn.tudelft.nl/users/bociort/networks.html 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.