Looking for order in the optical design landscape

Florian Bociort and Maarten van Turnhout
Optics Research Group, Delft University of Technology
Lorentzweg 1, 2628 CJ Delft, The Netherlands

ABSTRACT

In present-day optical system design, it is tacitly assumed that local minima are points in the merit function landscape without relationships between them. We will show however that there is a certain degree of order in the design landscape and that this order is best observed when we change the dimensionality of the optimization problem and when we consider not only local minima, but saddle points as well. We have developed earlier a computational method for detecting saddle points numerically, and a method, then applicable only in a special case, for constructing saddle points by adding lenses to systems that are local minima. The saddle point construction method will be generalized here and we will show how, by performing a succession of one-dimensional calculations, many local minima of a given global search can be systematically obtained from the set of local minima corresponding to systems with fewer lenses. As a simple example, the results of the Cooke triplet global search will be analyzed. In this case, the vast majority of the saddle points found by our saddle point detection software can in fact be obtained in a much simpler way by saddle point construction, starting from doublet local minima.

Keywords: saddle point, optimization, optical system design, lithography

1. INTRODUCTION

One of the main difficulties in optical system design is caused by the presence of multiple local minima in the merit function landscape. Since there are typically many minima, the solution that will be obtained by locally optimizing a design is critically dependent on the choice of the initial configuration. Major recent progress in global optimization software alleviates the difficulty, but the computational efficiency or even the applicability of many of these algorithms still remains an issue, especially for complex design problems. Global optimization methods that are in use today in optical system design rely almost exclusively on generally applicable mathematical algorithms, because for the design landscape no specific optical properties are known, that could be used effectively for moving the system from one local minimum to another. In this paper we show that a certain degree of order can be found in the optical design landscape and that this order is best observed when we consider not only local minima of the merit function, but saddle points as well. Such a-priori information about the design landscape can then be used to speed up the search for new local minima.

In this paper we will consider only those saddle points for which, intuitively, the merit function around them behaves like the two-dimensional surface of a horse saddle (the so-called Morse index 1 saddle points). Even when the number of optimization variables is large, such a point, if perturbed on one side of the "saddle", leads after optimization to one local minimum, and on the other side to a different minimum.

We have observed order at different levels. First, when certain quite general conditions are met, all minima and saddle points (of the type mentioned above) are linked in a network, such as the one shown in Fig.1 for the global search corresponding to the well-known Cooke triplet. For systems that are sufficiently simple, the entire network can be detected link by link, but this approach becomes difficult and time-consuming when the number of components increases. Fortunately however, certain properties, which can already be observed for simple systems, survive generalization and remain valid for systems of arbitrary complexity.
Figure 1. Network of the global search corresponding to the Cooke triplet \((m_1 \text{ and } m_2)\) for which the variables are the surface curvatures. The results are collected from a set of runs with different specifications (runs with object at infinity, runs for symmetric problems with transverse magnification -1, the field varies between 20 and 33 degrees). Individual runs may show minor departures from this figure, e.g. a pair consisting of a minimum and a neighboring saddle point may disappear or such an extra pair may appear. Since the network is symmetric, pairs of systems, which look like mirror images of each other, are grouped together for clarity in the same box. The minima are drawn within thick-line boxes and the saddle points within thin-line boxes. The two downward paths of local optimization started on both sides of the saddle at a saddle point \(s_{ij}\) lead to minima \(m_i\) and \(m_j\) as indicated by the continuous lines between systems. The dashed lines indicate a possible instability of the link when specifications are changed. Then, on one side of the saddle point \(s_{ij}\), the downward path may lead either to minimum \(m_i\) or to \(m_j\). ("\(\vee\)" denotes the Logical OR.) Extra information in this figure will be explained in Sec. 3.
In most cases, in global optimization the dimensionality of the problem (i.e. the number of optimization variables) is kept fixed. However, interesting relationships between local minima and saddle points can be observed in runs with different dimensionalities, and based on them practical design techniques can be developed. For instance, starting from a system which is a minimum with \( N \) optimization variables, a saddle point with \( N+2 \) optimization variables can be constructed by inserting in the system a lens in a certain way, as shown below. Lens designers often insert or split lenses in existing designs, but this is done in a manner that leads after optimization to a single local minimum. If, however, the lens is inserted in such a way that a saddle point is created, then two local minima result after optimization (one on each side of the "saddle") and the best of the two can be used for further design. If desired, the newly inserted lens or another lens in the system can then be removed.

Recently we have shown how saddle points can be constructed in the special case when the inserted lens is in direct contact with one of the surfaces of the original local minimum (the reference surface) and when the glass of the new lens is the same as that at the reference surface\(^{11}\). The theory behind the method contains concepts which are new in lens design, but the technique is very straightforward, can be easily integrated with traditional design techniques, and is applicable to systems of arbitrary complexity. In fact, our present experience shows that in the design of very complex systems such as lithographic objectives, for which other methods to move from a local minimum to another one are cumbersome, saddle point construction increases design productivity substantially and often reveals unexpected potential for further improvement of existing designs. It is for instance possible to remove up to three lenses from state-of-the-art lithographic objectives with high NA, without negative effects for their performance\(^{12}\).

In the next section, the saddle point construction method will be generalized so that the restrictions mentioned above can be removed. However, because of the novelty of the structure discovered in the design landscape, in Section 3 the emphasis will be on understanding the relationships between minima and saddle points when the dimensionality of the optimization problem is changed, rather than on practical applications.

### 2. GENERALIZING THE SADDLE-POINT CONSTRUCTION METHOD

We consider an optical system that is a local minimum with \( N \) optimization variables (called in what follows the "old" variables). We want to insert a lens in the system in such a way that we create a saddle point. The two new surfaces will have the indices \( k+1 \) and \( k+2 \) in the system. For simplicity, we consider the case when the lens surfaces are spherical, with the curvatures \( c_{k+1} \) and \( c_{k+2} \) as two "new" optimization variables, although the method can be easily generalized to aspherical surfaces. We have shown earlier\(^{11}\) that in the special case when the new lens is introduced in contact with a surface (with index \( k \)) of an already existing lens with the same glass, a saddle point is created if, for instance, the new lens has zero thickness and equal curvatures having the values

\[
c_{k+1} = c_{k+2} = c_k
\]

and all \( N \) old variables are left unchanged. Since a zero-thickness meniscus with equal spherical surface curvatures \( c_{k+1} = c_{k+2} \) disappears physically, we call it a "null" lens. (In addition to the saddle points given by Eq.(1), which we call the "concentric" saddle points, other saddle points with null lenses can also exist. The common curvature value can then be computed with the numerical method discussed below.)

We consider now the case when a null lens is inserted at an arbitrary position in the system and has an arbitrary glass. The merit function is an arbitrary optical merit function \( f \). Since a null lens does not affect the path of the light rays, \( f \) remains unchanged and equal to the merit function of the original local minimum with \( N \) variables for any value of the null-lens curvature, i.e. along an entire straight line in the variable space

\[
c_{k+1} = c_{k+2} = u, \quad \text{u variable, the old variables kept constant.}
\]

With a null lens, the partial derivatives of the merit function with respect to the old optimization variables remain unchanged, i.e. equal to zero. Saddle points are critical points, which means that for obtaining a saddle point the partial derivatives with respect to the new variables \( \partial f / \partial c_{k+1} \) and \( \partial f / \partial c_{k+2} \) must also vanish. Since along the line (2) we have
\[
d\frac{f}{dc_{k+1}} = \frac{\partial f}{\partial c_{k+1}} + \frac{\partial f}{\partial c_{k+2}} dc_{k+2} = 0
\]  

(3)

and \( c_{k+1} = c_{k+2} \), it follows that

\[
\frac{\partial f}{\partial c_{k+1}} = -\frac{\partial f}{\partial c_{k+2}}
\]

(4)

which means that both new components of the gradient vanish simultaneously. Thus, for finding the saddle points, we find numerically the values of the null-lens curvature for which

\[
\frac{\partial f}{\partial c_{k+1}} = 0.
\]

(5)

The critical points that are solutions of Eq. (5) are saddle points (and have therefore a direction along which \( f \) can decrease) because they cannot be local minima. For local minima in the variable space, the equimagnitude hypersurfaces of \( f \) around them are ellipsoids that reduce to a single point when \( f \) has the value that corresponds to the minimum. However, solutions of Eq. (5) are situated on equimagnitude hypersurfaces that contain an entire straight line, given by Eq. (2), along which \( f \) remains constant. See also Ref. 11.

Note that, because of Eq. (4), for finding a null-lens saddle point only one equation with one unknown \( c_{k+1} \) must be numerically solved. This technique is considerably more efficient than detecting a saddle point in the optical design space without a-priori knowledge of optical system properties, as in the case of the saddle points shown in Fig. 1. When a saddle point is found, we take on the line given by Eq. (2) two points on opposite sides of the saddle, having

\[
c_{k+1} = c_{k+2} = c_s \pm \epsilon
\]

(6)

where \( \epsilon \) indicates a small curvature change and \( c_s \) is a solution of Eq. (5). Then, these two points are optimized and two different local minima will result. Finally, in these solutions the zero distance between surfaces \( k+1 \) and \( k+2 \) can be increased to the desired values.

We have implemented the generalized saddle point construction method in the macro language of the optical design program CODE V. In all examples shown in this paper, the merit function is this program’s default error function based on the transverse ray aberration computed with respect to the chief ray. A simple example is shown in Fig. 2. For a doublet local minimum, in which a null lens has been inserted after the first lens, \( \frac{\partial f}{\partial c_3} \) is plotted as a function of \( c_3 \) (\( =c_4 \)). The curve has a parabolic shape and cuts the axis in two points.

In this example, the reason why Eq. (5) has two solutions can be easily understood by examining the drawings of the systems corresponding to the two saddle points. For s10-7 and s4-10, the null-lens curvature is almost the same as \( c_2 \) and \( c_5 \), respectively, despite the fact that the null lens is placed at some distance from the neighboring surfaces and that its glass differs from the glasses of the first and last lens. In fact, if for s10-7 the distance between the second and third surface is gradually decreased to zero, and the glass of the null lens is gradually changed into the glass of the first lens, then the saddle point becomes concentric and \( c_3 = c_4 = c_2 \) holds rigorously. Similarly, s4-10 can be continuously transformed into a rigorously concentric saddle point with \( c_3 = c_4 = c_5 \).

From the two null-lens saddle points, three local minima m4, m7 and m10 are obtained by using Eq.(6) and optimizing \( (m10 \) results from both saddle points). When the zero lens thickness is increased, the three minima become after reoptimization the systems indicated by black arrows. The same three nonzero thickness minima (m4, m7 and m10) and two nonzero thickness saddle points (s10-7 and s4-10) linked to them can also be obtained with our network-detection software. By comparing the null-lens saddle points with the corresponding nonzero thickness saddle points, we observe that with the exception of the thickness of the middle lens the systems are essentially identical. Thus, the fact that one lens in the saddle points constructed with this method has zero thickness is not an obstacle for theoretical analysis or practical applications, because from null-lens saddle points the same minima with nonzero thickness can always be obtained as from the corresponding nonzero thickness saddle points. It is just technically easier to increase the thickness in a local minimum than in a saddle point.
Figure 2. Construction of a triplet saddle point by inserting a null lens into a doublet local minimum with variable curvatures. The aperture, field, and wavelength specifications, the glass types and constant distances between surfaces are those for a Cooke triplet global search with object at infinity. The systems in the dashed box can be found in Fig. 1, their counterparts with a zero-thickness lens are underlined.

Figure 3. Minima in the achromatic doublet global search: G(Gauss), S(Steinheil), F(Fraunhofer), R(Reversed Gauss) and H(Hub).

Figure 4. The networks of local minima and saddle points for an air-spaced doublet. On the left: flint-crown glass order; on the right: crown-flint glass order. The "m" systems are local minima, and the "s" systems are saddle points with Morse index 1. The CODE V error function value is also given.

Figure 5. A minimum from Fig. 3 is changing shape when specifications are changed gradually in several steps.
3. WHY ARE THERE SO MANY MINIMA IN THE COOKE TRIPLET GLOBAL SEARCH?

In global optimization, the focus is usually on the quality, given by the value of their merit function, of the systems that correspond to different local minima. We believe, however, that the existence of local minima and their quality are distinct issues to be investigated. For instance, the network structure (i.e. the local minima, the saddle points, and the links between them) are often remarkably stable when specifications change, but the best local minima for some specifications do not necessarily remain the best ones for other specifications. As a simple example, we compare the networks corresponding to an achromatic air-spaced doublet global search in two different situations: the situation when the first lens has a flint glass and the second lens a crown glass, and the situation when the glass order is reversed. For thin achromatic doublets, for which the optimization variables are the surface curvatures, it is well known that there are four doublet shapes (G, S, F and R, see Fig.3.) for which the important aberrations (spherical aberration, coma and axial color) can be simultaneously corrected\(^{13}\). Another doublet minimum, H, also exists, but has poor imaging qualities, because not even spherical aberration is corrected\(^{14}\). For axial color correction, the crown glass element must be the one with positive power, so with crown glass first the best solutions are F and G and with flint first R and S are best, as can be also observed in Fig.4. Remarkably however, the two networks are identical, i.e. all five minima exist in both cases and are linked in the same way via the corresponding saddle points, despite the fact that the solutions which have good quality are not the same.

In general, when we change control parameters (i.e. system specifications and parameters that remain constant during optimization), the network structure remains unchanged until a control parameter reaches a certain threshold value. At threshold, local minima and saddle points appear or disappear in the networks. For many network links, the parameter ranges between threshold values are large, and therefore, despite significant changes of the control parameters, the corresponding local minima and saddle points continue to exist, with usually only minor modifications in the shape of the systems. For example, the doublet minima (especially G, S, F and R) are remarkably robust. In most cases, they continue to exist as local minima when the aperture and field specifications, as well as the distances between surfaces and the glass types, are gradually changed to Cooke triplet values, although the merit function deteriorates severely. (See Fig.5.) Also, the concentric saddle points, the null-lens saddle points constructed as shown in the previous section, and the local minima that are linked to them and still have zero distances between certain surfaces, continue to exist in most cases when the zero distances between surfaces are increased to values that are not too large.

Since we believe that understanding stable network structures is important, we ignore for the moment the imaging quality of the various minima and focus in what follows on why multiple local minima exist at all and on how they can be obtained in the given problem. If in a one-dimensional optimization problem a single minimum was expected, but two local minima were observed, the reason for this would become clear if one could understand why a maximum exists between the minima. In a multi-dimensional optimization problem, a saddle point with Morse index 1 is a maximum in one direction in the variable space, and a minimum in all other orthogonal directions. Since minima can be viewed as resulting from saddle points by letting the optimization roll down on both sides of the "saddle", in order to understand why multiple local minima exist, it is sufficient to understand why saddle points exist.

According to this line of reasoning, the presence of minima such as m4, m7 and m10 in Fig. 2 (which can be detected not only with our network search, but with other methods, such as Global Synthesis of CODE V, as well) can be viewed as a consequence of the presence of saddle points such as s10-7 and s4-10 which are linked to them. These saddle points can be viewed as resulting from null-lens saddle points (s10-7 and s4-10) which in turn are the two solutions of Eq.(5), applied to a doublet local minimum plus a null lens that has been inserted between the doublet lenses. In this way, the presence of a minimum in a doublet merit function landscape leads to the presence of saddle points (and of minima resulting from them via optimization) in a triplet landscape.

For studying the relationship between minima with \(N\) variables and saddle points with \(N+2\) variables, it is useful to examine simple situations first, in which the entire set of saddle points can be found, and see how many of them result from local minima with fewer lenses. (The corresponding minimum with \(N\) variables will be called a precursor of the saddle point with \(N+2\) variables and of the two minima linked to it.) In the doublet networks shown in Fig.4, note that all saddle points have the shape of a positive lens plus a meniscus. We also obtained all of them from singlets in a way similar to that shown in Fig. 2. We return now to the Cooke triplet global search shown in Fig. 1, which was obtained with our network detection software. It turns out that out of the 26 saddle points, 21 can be obtained from doublet local
minima plus a null lens, 19 of them in the way described in the previous section, and 2 with a more elaborate - but still one dimensional - calculation, as shown below. The exceptions are the 5 saddle points shown in gray boxes, which we could not obtain from doublets with a null lens. (In some sense they seem less important because they do not lead to interesting new local minima.)

Three sets of doublet local minima have been obtained by removing one of the Cooke triplet lenses (the first, or the second, or the third) and globally optimizing the remaining doublet with variable curvatures for Cooke triplet specifications. All doublet minima belong to the types shown in Fig. 3. With our generalized saddle point construction method, 19 triplet saddle points (those that have a hatched lens in Fig. 1) result from the doublets of types G, S, F, and R. By ignoring the hatched lens in the saddle point drawings, we can easily recognize the doublet type indicated in the figures. (Keep in mind that because of different specifications, the systems in Fig. 3 change shape as shown in Fig. 5.) In all cases, the hatched lens is a meniscus with almost equal curvatures. For zero thickness this lens becomes the null lens used in the saddle point construction method, as shown in Fig. 2. In a few drawings, a thin lens is drawn instead of a hatched one, indicating that the corresponding saddle point exists only for small thickness values, and that the saddle point disappears in some of the network runs where the thickness of that lens happens to be larger than the threshold value. One saddle point, s19-18, has two hatched lenses, which shows that it can be obtained in two different ways from doublets.

Two saddle points that lead to the two Cooke triplets (m₁ and m₂) result from a H-type doublet. However, in both cases, the extra lens (shown in black) has unequal curvatures. For the curvatures of the null lens placed between the two lenses of the H-type doublet, Eq. (5) does not have real solutions, i.e. the parabola-like curve in Fig. 2 does not cut the axis. However, by starting from the H doublet plus null lens and performing a one-dimensional succession of local optimizations along the line (2) with values of \( u \) close to the extremum of the parabola, we were able to reach m₁ and m₂ and also the local minima on the other side of s₂-(6v3) and s₁-(5v7). Remarkably, since these two saddle points have Morse index 1, they repel the solution in one direction (the downward direction of the “saddle”), but attract the solution in all other directions. Therefore, the optimization first converges toward them and then diverges, reaching local minima.

If the optimization is stopped in early stages, zero-thickness systems close to saddle points are obtained that resemble s₂-(6v3) and s₁-(5v7), as shown in Fig. 6.

Figure 6. For certain starting values of \( c_i = c_4 = u \), the optimization algorithm leads after a certain number of cycles to systems that strongly resemble the saddle points s₂-(6v3) and s₁-(5v7). For both saddle points, the left drawing shows the thin-lens intermediate optimization result (underlined). The right drawing actually shows the superposition of two drawings: the exact saddle point found with our network search, and the drawing on the left in which the thickness of the middle lens has been increased to facilitate comparison. Note that for both saddle points the two drawings on the right side are almost indistinguishable, indicating that the optimization has arrived very close to the real saddle points.

Figure 1 was obtained with our network search algorithm, which is generally applicable to any optimization problem with continuous variables and does not use information about specific optical system properties. We observe however that the vast majority of the triplet saddle points and local minima obtained with our network search can be obtained in a much simpler and faster way by using the saddle point construction method, which does use such extra information.
Finally, by examining how systems that have the same precursor are arranged in the Cooke triplet network, we observe an interesting structure, shown in Fig. 7. The systems (minima and saddle points) that have the same precursor turn out to be all connected to each other by network links. With the exception of the few saddle points in gray and of local minima linked only to these saddle points, the entire network is partitioned in regions in which all systems result from the same doublet type F, S, G, R or H. (Because of symmetry, Fig. 7 shows three regions: F-S, G-R and a small H region.) In the F, S, G, and R regions, all saddle points -except one in each region- are linked to the same minimum, called the "hub" of that region. In fact, in the doublet networks in Fig. 4 we encounter a similar feature: all doublet saddle points -they can be obtained from the same singlet- are linked to the same minimum, the doublet hub H. For certain specifications, hubs that have hubs as precursors can be found, e.g. the Fulcher system\textsuperscript{11}. Such "super" hubs have, we believe, interesting properties.

The doublet regions in the Cooke triplet network overlap at certain local minima (which we call linking minima) thus ensuring the connectivity of the entire network. Because, as shown in Fig.1, a few network links (drawn dashed) are unstable, for some specifications instead of the main linking minimum (shown with overlapping hatching) another minimum (in brackets) takes over the role of linking minimum. The systems m\textsubscript{3} and m\textsubscript{6} in Fig. 1, which in Fig. 7 are shown as linking minima for three different doublet regions, are particularly interesting. By adding lenses to them via saddle point construction, we could obtain the Double Gauss design.

Figure 7. Partition of the Cooke triplet network in distinct connected regions, such that the systems in each region have the same doublet precursor. The systems in the boxes (drawings not shown) are the same as in Fig. 1. Different regions are marked with different hatching and are separated by thick continuous lines. Several local minima that have more than one precursor form links between distinct regions. Linking minima are shown in thick dashed boxes in which the hatching is an overlap of the hatching of the regions they link.
4. CONCLUSIONS

In the complex topography of the optical merit function landscape a surprising "hidden" structure has been found. This structure is best observed when we compare the results of global searches with different dimensionality and when we consider not only local minima, but saddle points as well. We have shown that for simple systems, many saddle points result from adding null lenses to minima with a lower dimensionality. Thinking in terms of saddle points is still unfamiliar to most optical system designers, but we believe that the potential for discovering new solutions and for improving design productivity justifies the effort for understanding such new methods and for combining them with traditional design methodology.

ACKNOWLEDGMENTS

We thank Chiung-tze (Joanne) Wang and Alexander Serebriakov for their contributions to the data in Figures 1 and 4, respectively. We also acknowledge the use for this work of an educational license of CODE V. The work of the second author was supported by the Dutch Technology Foundation STW.

REFERENCES

10. More examples of networks can be found in papers, which are also available at http://wwwoptica.tn.tudelft.nl/users/bociort/networks.html
11. F. Bociort and M. van Turnhout, "Generating saddle points in the merit function landscape of optical systems", Proc. SPIE 5962, 0S1-8, 2005