

Instabilities and fractal basins of attraction in optical system optimization

M. van Turnhout and F. Bociort

Department of Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

mvanturnhout@gmail.com

Abstract: Many optical design programs use various forms of the damped least-squares method for local optimization. In this paper, we show that damped least-squares algorithms, with maximized computational speed, can create sensitivity with respect to changes in initial conditions. In such cases, starting points, which are very close to each other, lead to different local minima after optimization. Computations of the fractal capacity dimension show that sets of these starting points, which lead to the same minimum (the basins of attraction for that minimum), have a fractal structure. Introducing more damping makes the optimization process stable.

© 2008 Optical Society of America

OCIS codes: (080.1753) Computation methods; (080.2720) Mathematical methods (general); (220.2740) Geometric optical design; (220.3620) Lens system design.

References and links

1. D. C. Sinclair, "Optical design software," in *Handbook of Optics, Fundamentals, Techniques, and Design, Vol. 1, 2nd ed.*, M. Bass, E. W. Van Stryland, D. R. Williams, and W. L. Wolfe, eds. (McGraw-Hill, New York, 1995), 34.1–34.26.
2. F. Bociort, "Optical system optimization," in *Encyclopedia of optical engineering*, R. G. Driggers, ed. (Marcel Dekker, New York, 2003), 1843–1850.
3. D. P. Feder, "Automatic optical design," *Appl. Opt.* **2**, 1209–1226 (1963).
4. E. Ott, *Chaos in dynamical systems, 2nd ed.* (Cambridge University Press, Cambridge, 2002).
5. H. Gross, H. Zügge, M. Peschka, and F. Blechinger, "Principles of optimization," in *Handbook of Optical Systems, Vol. 3* (Wiley-VCH, Weinheim, 2007), 291–370.
6. C. Grebogi, S. W. McDonald, E. Ott, and J. A. Yorke, "Final state sensitivity: an obstruction to predictability," *Phys. Lett. A* **99**, 415–418 (1983).
7. S. W. McDonald, C. Grebogi, E. Ott, and J. A. Yorke, "Fractal basin boundaries," *Physica D* **17**, 125–153 (1985).
8. C. Grebogi, E. Kostelich, E. Ott, and J. A. Yorke, "Multi-dimensioned intertwined basin boundaries: basin structure of the kicked double rotor," *Physica D* **25**, 347–360 (1987).
9. C. Grebogi, E. Ott, and J. A. Yorke, "Chaos, strange attractors, and fractal basin boundaries in non-linear dynamics," *Science* **238**, 632–638 (1987).
10. H. E. Nusse, and J. A. Yorke, "Basins of attraction," *Science* **271**, 1376–1380 (1996).
11. M. van Turnhout, and F. Bociort, "Predictability and unpredictability in optical system optimization," in *Current Developments in Lens Design and Optical Engineering VIII*, P. Z. Mouroulis, W. J. Smith, and R. B. Johnson, eds., Proc. SPIE **6667**, 666709 (2007).
12. Optical Research Associates, CODE V, Pasadena, CA.
13. ZEMAX Development Corporation, ZEMAX, Bellevue, WA.
14. F. Bociort, E. van Driel, and A. Serebriakov, "Networks of local minima in optical system optimization," *Opt. Lett.* **29**, 189–191 (2004).
15. F. Bociort and M. van Turnhout, "Generating saddle points in the merit function landscape of optical systems," in *Optical Design and Engineering II*, L. Mazuray and R. Wartmann, eds., Proc. SPIE **5962**, 59620S (2005).
16. S. N. Rasband, *Chaotic dynamics of nonlinear systems* (Wiley, New York, 1990).
17. H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and fractals: New frontiers of science, 2nd ed.* (Springer-Verlag, New York, 2004).

1. Introduction

In the field of optical system design, the aim of optimization is to find a system configuration that fulfills certain design targets within tolerances determined by the application. Typically, there are more targets than optimization variables, and only a least-squares solution can be found, where all targets are met collectively as close as possible. The merit function of the problem, which expresses the quality of an optical system, must therefore be minimized [1, 2].

When we represent a lens configuration having N optimization variables as a single point in an N -dimensional space of system parameters, the optimization process consists of moving from one point to another, reducing the merit function after each step, until a minimum is reached. Close enough to a minimum, further iterations will not produce any significant changes in the system parameters [3]. However, the solution that will be obtained after local optimization is critically dependent on the choice of the initial configuration. In general, the merit function is highly nonlinear. This typically gives rise to many local minima in the merit function landscape.

In order to determine the initial configurations that lead to a given local minimum, the so-called basin of attraction for that minimum should be considered. The set of all starting configurations that are attracted to a local minimum is the basin of attraction for that minimum [4] (see Fig. 1). A plot of the basins for all starting points after a finite number of iterations can be interesting for optical system design, because the basins show the sensitivity of the optimization result with respect to the starting values.

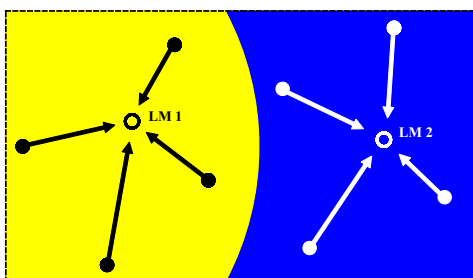


Fig. 1. Two neighboring basins of attraction. After local optimization, all initial points in one basin lead to the same local minimum (LM1 or LM2).

The damped least-squares method is particularly suitable to solve nonlinear least-squares problems, in which the merit function is the weighted sum of squares of operands that describe the design targets. Many optical design programs use damped least-squares methods for local optimization in various forms with great success [1, 5]. In damped least-squares methods, the operands are considered to be linear in a first approximation; the damping factor was originally introduced with the purpose of limiting the change of the variables to ensure that this linear approximation remains valid. However, in optimization programs, the damping factor is frequently chosen to increase computational speed. For example, it could be chosen such that the merit function decrease per optimization step is maximized.

In this paper, we show that choosing low damping factors for the sake of increasing computational speed can create sensitivity with respect to initial conditions. In these cases, the basins of attraction have very complicated structures with non-integer fractal dimensions, similar to those encountered in research on dynamical systems [4, 6, 7, 8, 9, 10]. One of the goals of this paper is to make optical system designers aware of the possibility of presence of such instabilities in the optimization process, because they can influence the result that will be obtained after optimization. In a more provisional form, we have addressed this type of behavior in Ref. [11].

It is well-known that a change in the optimization strategy or in the settings of the local

optimization algorithm can lead to a different result, even when starting from the same configuration [5]. When the optimization results are distinct for different optimization methods, then the basins of attraction obtained with these methods are also different. We observe this property when we use different damped least-squares software implementations to obtain the basins of attraction for the same optimization problem.

Since computing the basins of attraction is very time-consuming, we use a simple lens configuration to illustrate the appearance of fractal behavior. A monochromatic doublet with only two variables will be the main example of this paper. However, we also show that fractal behavior can be found in practical situations with more than two variables as well.

To have maximum stability, we first use a differential equation to go down the gradient of the merit function (Sec. 2). The basins of attraction obtained with this method are used as a reference for the basins obtained with different programs using some form of a damped least-squares method (Sec. 3).

In order to explain the peculiar behavior we observed in commercially available codes, where we have no control over the damping, we studied this behavior in more detail with the program OPTSYS, written by Joseph Braat. This program gives us more control over the damping. We will demonstrate that the choice of the damping factor is essential in obtaining stable results.

To further understand the fractal basin shapes, we examine the steps taken by the optimization algorithms of two commercially available programs in the variable space of the doublet example (Sec. 4). In addition, we demonstrate a procedure to change the step sizes in such a way that the basin boundaries become regular. With regular basin boundaries, the instabilities in the optimization process disappear.

2. Method for stable behavior

One technique used to find minima in the merit function (MF) landscape is to continuously follow the gradient of the merit function downwards, by solving the differential equation:

$$\frac{d\mathbf{x}}{ds} = -\frac{\nabla\text{MF}}{\|\nabla\text{MF}\|}, \quad (1)$$

where the vector $\mathbf{x} = (x_1, x_2, \dots, x_N)$ describes a point in the N -dimensional solution space, and $ds = \sqrt{\|d\mathbf{x}\|^2}$ is the arc length along the downward path. Although this method is rather slow, it gives maximum stability to the minimization process and does not display fractal behavior in our examples. The basins of attraction obtained with this method are used as a reference for the basins generated by three different optimization algorithms, all of which are based on some version of the damped least-squares method: OPTSYS, and two commercial codes CODE V [12] and ZEMAX [13]. All these programs adapt the damping factor in the damped least-squares method automatically, but OPTSYS allows the user to specify the maximum value of the damping factor to be used.

The main example in this paper is a monochromatic doublet (f number 3), with the curvature of the second and third surfaces (c_2 and c_3 , respectively) used as variables. All specifications are given in Tabs. 1 and 2. The first curvature of the doublet is kept constant, and the last curvature is used to keep the effective focal length constant at 100 mm. The stop is placed at the first surface. The image plane is kept at the paraxial image position.

Depending on the software used, optimizing the two variable curvatures results in four or five different local minima, which are shown in Fig. 2. Minima A–D are stable local minima, and minimum B has a high merit function value. For the settings given in Tabs. 1 and 2, minimum E (which also suffers from some edge thickness violation) is less stable, and appears or disappears easily when we change some parameter or the software. The control of edge thickness violation was disabled in this example, in order to eliminate the constraints from the list of possible

Table 1. Specifications for the monochromatic doublet example.

Entrance pupil diameter (mm)	33.333
Effective focal length (mm)	100.000
Paraxial image height (mm)	5.241
Field points (deg)	0.000, 2.000, and 3.000

Table 2. Curvatures, thicknesses, and refractive indices of the starting points for the monochromatic doublet example. For curvatures c_2 and c_3 the variation domain is given in brackets.

Surface #	Curvature (mm^{-1})	Thickness (mm)	Refractive index
object	0	infinity	1.
1	0.020	10.346	1.618
2	(-0.025, 0.040)	1.000	1.
3	(-0.045, 0.075)	2.351	1.717
4	solve	solve	1.
image	0	0.000	1.

causes of the peculiar behavior described in what follows. Detailed analysis shows no qualitative distinction in this respect between regions of the variable space where edge thickness constraints are satisfied or violated.

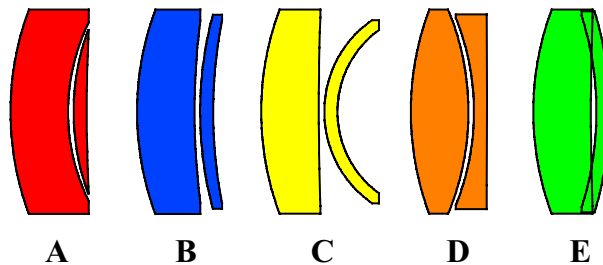


Fig. 2. Five local minima for a two-dimensional monochromatic doublet. The colors correspond to the colors used for the basins of attraction in subsequent figures.

In Fig. 3, we show the merit function equimagnitude contours (i.e. the contours along which the merit function remains constant) of the default merit function of CODE V, which is based on transverse ray aberrations (root-mean-square spot size) with respect to the chief ray. In the merit function, all fields have a weight factor of unity. In the same figure, we also plotted the positions of the minima A–D (large blue dots) and of the saddle points (small purple dots), which are situated at the crossing of two segments of the equimagnitude merit function contours [14, 15]. By taking the merit function of minimum B as reference, the relative merit function values for A, C and D are 0.01, 0.08 and 0.02, respectively. The three saddle points have relative merit function values close to 1.2. In this and following figures, curvature c_2 is plotted along the vertical axis, and curvature c_3 is plotted along the horizontal axis.

For the settings used in Fig. 3(a), minimum E did not exist. However, with slightly modified settings, minimum E can appear in CODE V as well. For example, when we decrease the glass thickness of the second lens to 0.3 mm, minimum E also appears in the merit function landscape made by CODE V.

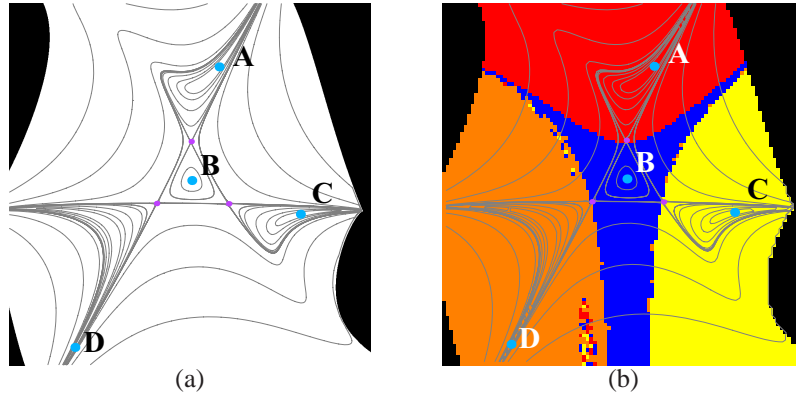


Fig. 3. (a) Merit function equimagnitude contours of the two-dimensional monochromatic doublet example (specifications are given in Tabs. 1 and 2). The larger blue points A, B, C, and D correspond to the local minima shown in Fig. 2. The smaller purple points are saddle points. In all figures of the same kind, the compact black regions contain configurations that suffer from ray failure. (b) Basins of attraction for the same doublet example (on a grid of 101×101 points), obtained by using differential equation (1) in a first stage of the optimization process in CODE V. The colors correspond to the local minima as shown in Fig. 2.

To compute the basins of attraction for the doublet, we first associate colors to the various minima that result from local optimization. We then specify a grid of equally spaced starting points in a two-dimensional plane (the variation domain is given in Tab. 2). At every grid point, we start to optimize the corresponding configuration until it arrives in one of the local minima. Depending on which local minimum we obtain, we color the starting point with the corresponding color for that minimum (see Fig. 2). Thus, the location of the point shows the initial value of the variables, and its color shows to which local minimum it converges to. Starting points for which the initial configuration suffers from ray failure (a ray misses a surface or total internal reflection) are shown in black.

To achieve maximum predictability of the downward path from the starting point to the local minimum, we first use differential equation (1). Using CODE V, we implement a fourth-order Runge-Kutta method with an adaptive step size in the macro language to solve Eq. (1). We determine the basins of attraction by choosing each point of a grid of 101×101 points as starting points. Since solving Eq. (1) is very time-consuming, we use a finite number of steps. In a second stage, we continue with the default local optimization routine of CODE V.

For the vast majority of grid points, after using Eq. (1), the starting point for the default optimization routine in the second stage is close enough to one of the local minima, and the damped-least-squares method becomes stable. Figure 3(b) shows the resulting basins of attraction for the doublet example. These will be our references for the following runs.

The basins shown in Fig. 3(b) are compact regions in the merit function space, and the boundaries between them are smooth curves. However, there are some minor artifacts, mainly along the basin boundaries, particularly prominent close to the upper and lower part of the boundary between the basins of minimum B and D. The artifacts are caused by the fact that we only use a finite number of steps with differential equation (1). When we increase the number of steps, and switch to default optimization closer to the minimum, the artifacts disappear. However, the cost for doing this is a considerable increase in computation time. Note that the saddle points in the merit function landscape are situated exactly on the basin boundaries.

3. Fractal basins of attraction

In this section, we demonstrate the fractal behavior of various forms of the damped least-squares method with plots of two-dimensional cuts of the basins of attraction. First, we discuss the basins of attraction of the two-dimensional monochromatic doublet, and then we show that we obtain similar behavior for a typical design problem with seven variables.

3.1. Two-dimensional monochromatic doublet

When we use the program OPTSYS with the settings given in Tabs. 1 and 2, we obtain all five doublet local minima which are shown in Fig. 2. Figure 4 shows the basins that are obtained with OPTSYS for two different values of the maximum damping factor. By changing the maximum damping factor, the shapes of the basins change as well. When the maximum damping is sufficiently large, we obtain the basins shown in Fig. 4(a), which look quite similar to the ones shown in Fig. 3(b).

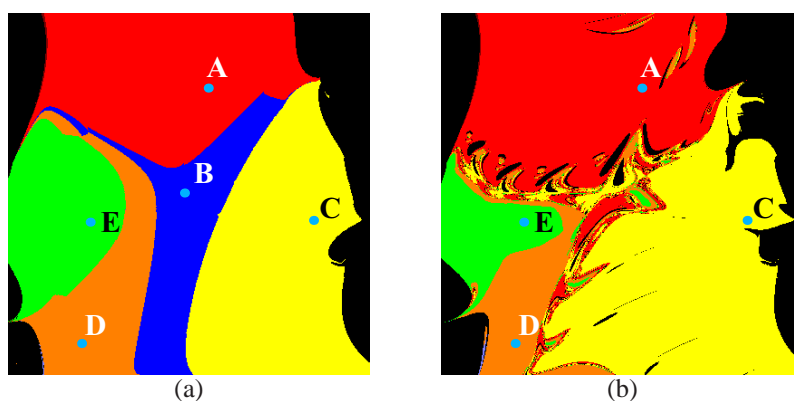


Fig. 4. Basins of attraction of the two-dimensional monochromatic doublet, obtained with OPTSYS, for different values of the maximum damping factor on a grid of 401×401 points. The specifications of the doublet are given in Tabs. 1 and 2. The colors correspond to the local minima as shown in Fig. 2. (a) Large damping, (b) 10 times smaller damping.

It is known that the behavior of the damped-least-squares optimization with high damping begins to resemble the behavior of the steepest descent method [5]. Therefore, it is not surprising that the basin boundaries in Fig. 4(a) are qualitatively of the same kind as the ones shown in Fig. 3(b). Note that the differential equation method used in Fig. 3(b) is essentially a steepest descent with infinitesimal step size.

By decreasing the damping, the regular basin shapes change into more complicated ones with increasingly diffuse boundaries. Since less damping allows the algorithm to take larger steps in the merit function landscape, the number of ray failures increases. When ray failure occurs during optimization, OPTSYS stops the optimization process. (CODE V and ZEMAX can seemingly escape from these situations.) The corresponding starting points are then drawn in black. As will be discussed in more detail in the next section, there are more black points in Fig. 4(b) (low damping) than in Fig. 4(a). When the maximum damping is too small, the algorithm does not converge to minimum B anymore, and the corresponding blue basin completely disappears [see Fig. 4(b)].

Several regions of basin boundaries from Fig. 4(b) are shown magnified in Fig. 5. When we examine these magnified figures, we notice the presence of very fine-scale interwoven structures. By using the concept of non-integer or fractal dimension [4, 9, 16], we can quantitatively

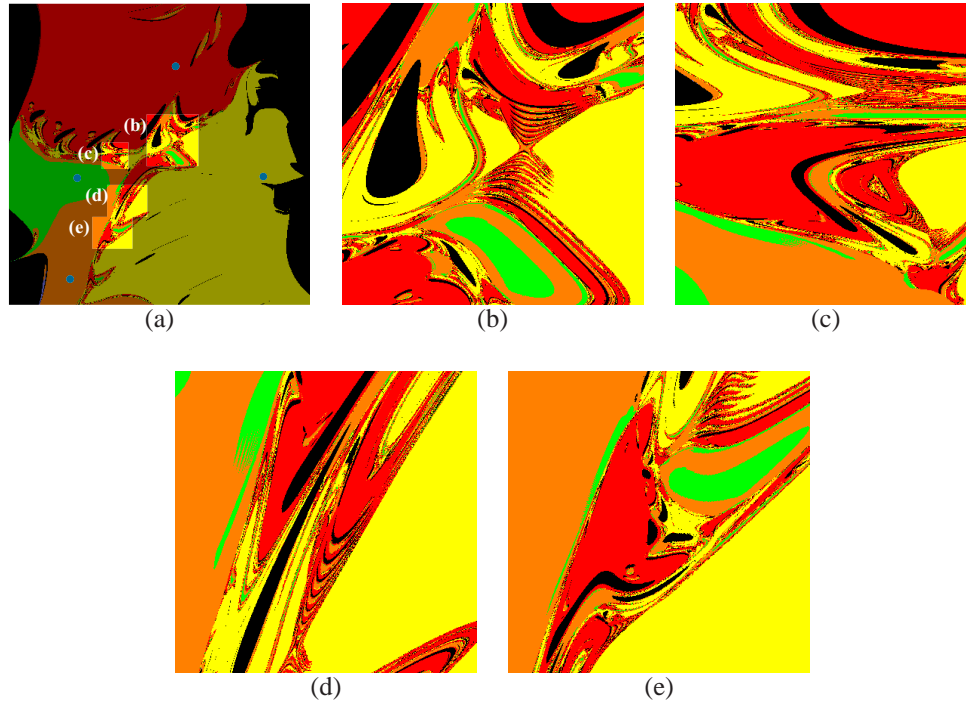


Fig. 5. Magnified regions of basin boundaries shown in Fig. 4(b). Figure 5(a) shows the positions of these regions within the parameter domain shown in Fig. 4(b).

investigate an important property of all basins, and show that the basin structure is fractal. There are several ways to compute fractal dimensions. In this paper, we use a simple method known as the box-counting method to calculate the so-called capacity dimension. A simple illustration of this method can also be found in Ref. [17].

In order to compute the fractal capacity dimension (D) for one basin, we cover the figure with a set of grids having mesh sizes ε , and count for each ε the number of grid boxes $N(\varepsilon)$, which contain at least one point of the considered basin. For small values of ε , the total number of grid boxes (N) needed to cover the basin is inversely proportional to ε^D :

$$N(\varepsilon) = k\varepsilon^{-D}, \quad (2)$$

where k is a constant. As ε is made smaller, the area of the considered basin covered by the grid boxes approaches the true area of that basin. The capacity dimension is defined as the value of D in Eq. (2), in the limit $\varepsilon \rightarrow 0$ [4, 9, 16].

The capacity dimension also includes the topological dimension, where, for a single point, $N(\varepsilon) = 1$ regardless of ε , giving $D = 0$ [16]. For a line with length L , we have $N(\varepsilon) = L/\varepsilon$, giving $D = 1$, and for a surface of area S , we obtain $N(\varepsilon) = S/\varepsilon^2$, which yields $D = 2$.

As an example of computing fractal dimensions for basins of attraction, Fig. 6 shows the orange basin shown in Fig. 5(b), covered with five different mesh sizes. It is convenient to use a sequence of grids where the initial mesh size ε_0 is the smallest, and the remaining mesh sizes increase by a factor of 2 from one grid to the next: $\varepsilon_0 = \varepsilon_1/2 = \varepsilon_2/4 = \dots = \varepsilon_m/2^m$, with $m = 0, 1, 2, 3, 4$, corresponding to grids with 400×400 , 200×200 , 100×100 , 50×50 , and 25×25 points, respectively. When a grid box contains at least one point of the orange basin, we color the complete box orange. The other basins are not considered and are colored white.

Let a be the length of the two sides of Fig. 5(b), then the mesh sizes are given by $a/400$, $a/200$, $a/100$, $a/50$ and $a/25$, respectively. We can rewrite Eq. (2) as:

$$\frac{N(\epsilon_m)}{N(\epsilon_0)} = \left(\frac{\epsilon_m}{\epsilon_0}\right)^{-D} = \left(\frac{1}{2}\right)^{mD}, \quad (3)$$

where the constant k in Eq. (2) is now given by $k = N(\epsilon_0)\epsilon_0^D$. Solving Eq. (3) for mD gives us:

$$mD = \log_2 \frac{N(\epsilon_0)}{N(\epsilon_m)}. \quad (4)$$

Equation (4) shows that when we plot the base 2 logarithms of $N(\epsilon_0)/N(\epsilon_m)$ versus m , then the slope of the line that fits the data gives us an estimation for D . For all fractals studied in this paper, the data points $(m, \log_2 N(\epsilon_0)/N(\epsilon_m))$ turn out to be very close to a straight line. After fitting the data for the orange basin, we obtain an almost straight line with slope 1.48. The data with linear fit are shown in Fig. 7.

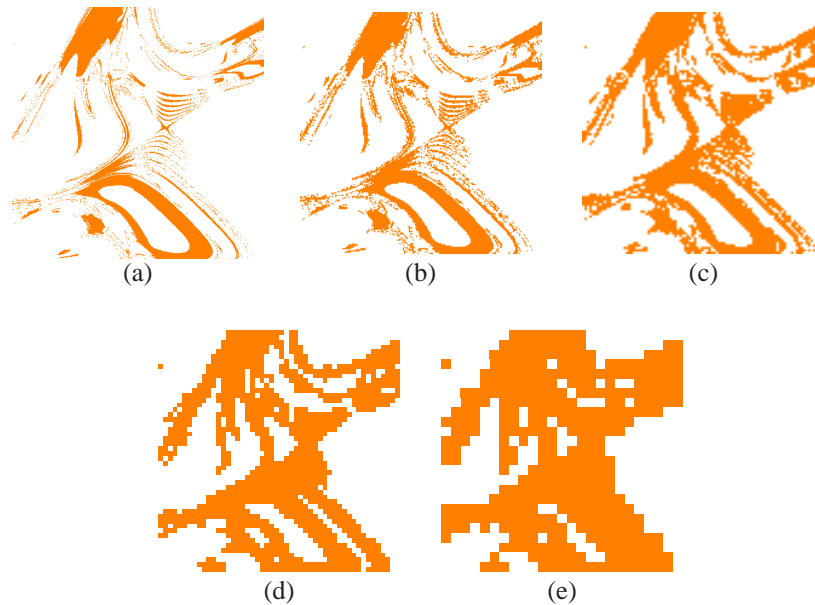


Fig. 6. Grid boxes that contain at least one point of the orange basin shown in Fig. 5(b), for five different grids with (a) 400×400 , (b) 200×200 , (c) 100×100 , (d) 50×50 , and (e) 25×25 points. The values of $N(\epsilon)$ are 23981, 8734, 3328, 1149, and 375, respectively.

Using Eq. (4), we calculate the capacity dimensions D of all remaining basin structures shown in Fig. 5 (listed in Tab. 3). The dimensions vary between $D = 1.16$ and $D = 1.87$. The fact that D is not an integer (or very close to an integer) demonstrates that the basin boundaries are fractal [4, 9].

Note that, for a given local minimum, the value of D for the entire figure is different from the value of D for the individual parts of the figure. This shows that the exact value of the fractal dimension also depends on the specific choice of the variation domain of the starting point parameters (c_2 and c_3 in this case). Therefore, the emphasis in this paper is on the fact that the fractal dimension of the basins we examine is not an integer, rather than on the specific value of the fractal dimension in these cases. The fractal nature of the basins shows that the same

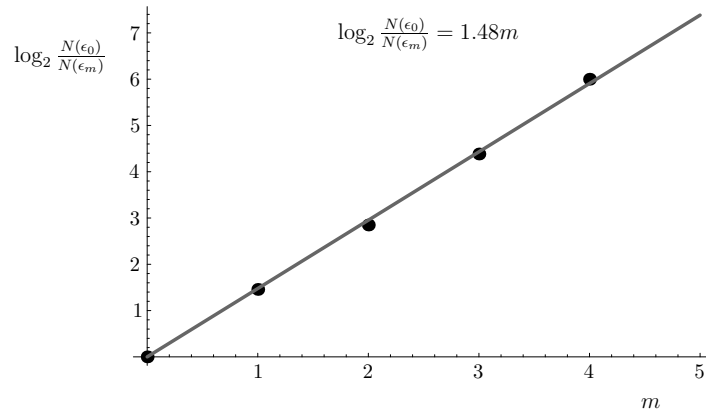


Fig. 7. Calculation of the capacity dimension D for the orange basin shown in Fig. 5(b). The slope of the straight line, which fits the data, gives $D = 1.48$ [see Eq. (4)].

complex mixture of basins shown in Fig. 5 is also present on arbitrarily small scales. The line in Fig. 7 can be extrapolated to the left as much as desired.

Table 3. Capacity dimensions of the basins of attraction shown in Figs. 5(a)–(e). The letters A, C, D, and E correspond to the basins of the corresponding local minima, and ‘RF’ refers to starting configurations for which ray failure occurs during optimization. Recall the basin of minimum B has disappeared in Fig. 5.

	5(a)	5(b)	5(c)	5(d)	5(e)
A	1.86	1.65	1.74	1.65	1.69
C	1.87	1.75	1.66	1.83	1.81
D	1.60	1.48	1.60	1.66	1.77
E	1.56	1.25	1.16	1.16	1.32
RF	1.71	1.51	1.35	1.34	1.21

For the same doublet problem, we also examined the basin boundaries obtained with commercial optical design software. We used CODE V and ZEMAX with their default damped-least-squares optimization algorithms and merit functions. In CODE V, for each initial set of curvatures, the system was iterated with a maximum of 100 optimization cycles (in ZEMAX with 30 cycles). For the settings we use here, minimum E (see Fig. 2) does not exist in the merit function landscape of CODE V or ZEMAX.

The resulting basins of attraction for the two-dimensional monochromatic doublet are shown in Fig. 8. In Figs. 8(a)–(d), we used CODE V on a grid of 1001×1001 points, and we used ZEMAX on a grid of 401×401 points in Fig. 8(e). The lens data of the doublet in Figs. 8(a)–(c) and (e) are given in Tabs. 1 and 2. In Fig. 8(d), the thickness of the second lens is changed to 0.3 mm. Figures 8(b) and (c) show magnified regions of Fig. 8(a), indicated with white rectangles, where we zoom in on some parts of the basin boundary between minimum A and D. Note the very complex shapes and the fine-scale structure in the basins of attraction, which is present on all scales. The basin boundaries show very finely interwoven basins, indicating the fractal nature of the boundaries. When we compare Figs. 8(a), (d), and (e) (obtained with two commercial programs), with Fig. 4(b) (obtained with OPTSYS), we observe certain similarities. However, since the two commercial programs can handle ray failure, if it occurs during optimization, better than OPTSYS, Figs. 8(a), (d) and (e) have less black points than Fig. 4.

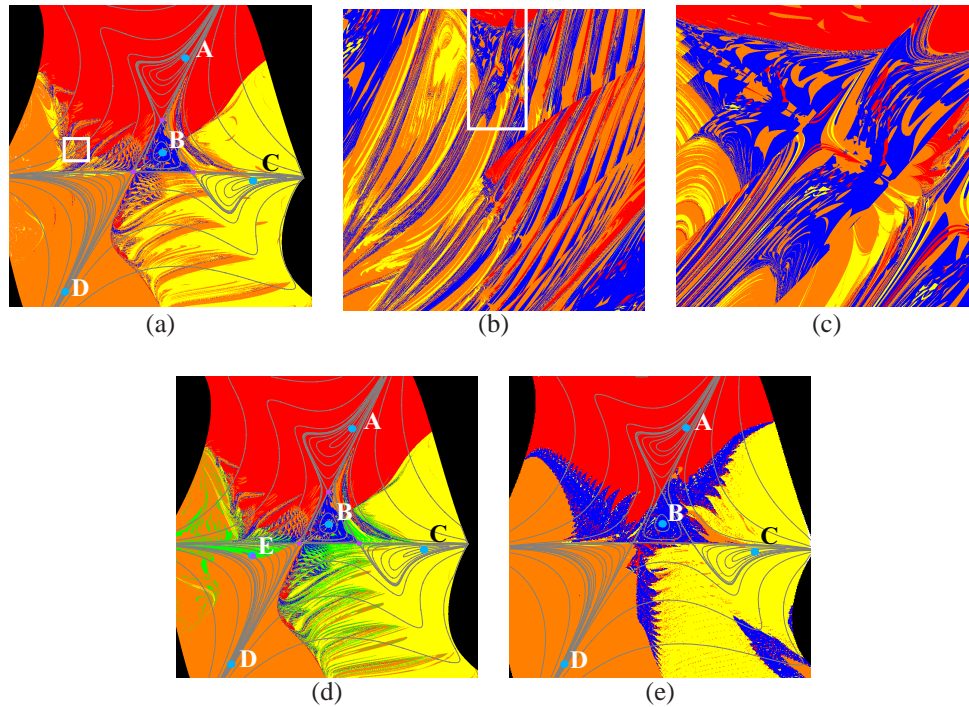


Fig. 8. Basins of attraction for the two-dimensional monochromatic doublet obtained with CODE V (a–d) and Zemax (e). (The gray contours are the equimagnitude contours of the merit function.) (a) Doublet with settings as given in Tabs. 1 and 2 and default optimization of CODE V. (b) Magnification of the small white rectangle in a). (c) Magnification of the small white rectangle in b). (d) Same as a), but with the thickness of the second lens equal to 0.3 mm. (e) Same as a), but with the damped-least-squares optimization of ZEMAX.

The capacity dimensions D of the basins of attraction shown in Fig. 8 are listed in Tab. 4. All basins have a capacity dimension between 1 and 2, indicating the fractal nature of the basin boundaries. In the regions where the compactness of a basin increases, its capacity dimension also increases.

Figure 8 shows a substantial difference between the basin shapes obtained with the algorithms implemented in the two commercial programs and our much slower algorithm based on Eq. (1). In this example, the finely interwoven fractal basins reveal the existence of extreme sensitivity to initial configurations in the default optimization routines of CODE V and ZEMAX, which causes unpredictable results in the optimization process. A small change in the starting configuration of a system, which is initially in a fractal region, can lead to a different local minimum after optimization.

3.2. Seven-dimensional optimization problem

The fractal behavior shown in the doublet example was also frequently found in basin plots for systems with more than two variables. As an example, we show the basins of attraction, obtained with CODE V, for a typical optimization problem with seven variables, where the goal is to design a Double Gauss system (without vignetting), see Fig. 9(a).

We use the default merit function of CODE V with all lens curvatures as optimization variables, except for the curvatures of the two plane cemented surfaces, which are kept unchanged,

Table 4. Capacity dimensions of the basins of attraction shown in Figs. 8(a)–(e). The letters A–E correspond to the basins of the local minima.

	8(a)	8(b)	8(c)	8(d)	8(e)
A	1.88	1.32	1.69	1.89	1.77
B	1.32	1.60	1.76	1.30	1.69
C	1.80	1.26	1.45	1.77	1.81
D	1.81	1.67	1.69	1.71	1.74
E	-	-	-	1.24	-

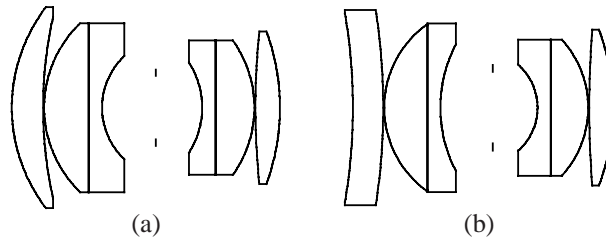


Fig. 9. (a) Double Gauss system (without vignetting). (b) Another local minimum which is found after optimizing a two-dimensional set of starting points (see text).

and the last curvature, which is used to keep the effective focal length constant.

In Fig. 10, we show a two-dimensional cut of the basins of attraction for this optimization problem. We choose a two-dimensional set of starting points, for which the curvatures of the first and second surfaces are varied within a certain range. The starting values for the other variables are identical for all these points.



Fig. 10. Basins of attraction for a seven-dimensional Double Gauss optimization problem. The initial curvature of the first surface is plotted vertically, and the initial curvature of the second surface is plotted horizontally.

We observe the presence of three local minima. In addition to the expected Double Gauss solution (the red basin in Fig. 10), we also obtain the minimum shown in Fig. 9(b) (the green basin) and a third minimum (the blue basin). The third minimum resembles the system shown in Fig. 9(b), except that the first lens has stronger curvatures. The Double Gauss design has a merit function value approximately 30 times smaller than the values for the other two systems.

Note that the basins of attraction for the minima again have irregular shapes. The fractal

dimensions computed for the red, green, and blue basins are: 1.87, 1.85, and 1.88, respectively. Although not shown in this work, we have computed fractal dimensions for basin boundaries of optimization problems having an even larger number of variables than in the examples shown here.

4. Instabilities in optimization

To understand why basin boundaries can be so complicated, we examine the sequence of steps in the two-dimensional variable space of the monochromatic doublet that are taken by the optimization routines of CODE V, OPTSYS, and ZEMAX. Examples of optimization paths are shown in Fig. 11. For observing details better, these figures can be enlarged on-screen in the electronic version of this paper. The starting points for optimization are indicated with the label 'START', and the result after each optimization cycle is shown in the figure by a dot. The dots are connected with lines, showing the direction of the optimization path in the merit function landscape.

In Fig. 11, we see a remarkable behavior of optimizations started in a fractal region. In all three examples, the four starting points are very close to each other, but they converge to four different local minima after optimization. When optimization is started in a fractal region, it first displays an unpredictable chaotic path in the merit function landscape, before it finally converges to a local minimum. For instance, in Fig. 11(a), the distances between the four points increase from cycle to cycle during the first cycles of optimization. After a number of iterations, the optimization starts to converge toward one of the four local minima. This type of behavior is known as a chaotic transient [4].

The iterations displayed in Fig. 11(b) show similar behavior. In fact, we encounter unexpected ray failure [black points in the middle of Fig. 11(b)], where the curvatures have moderate values. This is because the jumps are so large, that for some iterations the solution reaches the compact black regions (close to the borders of the figure) where ray failure occurs (due to large curvatures). However, the behavior shown in Fig. 11(c) is different. The four points are first attracted by the almost horizontal equimagnitude curve of the merit function in the middle of the figure. There, for each point, the optimization iterates a number of small steps close to this curve, until it is suddenly repelled by the curve, and converges to one of the four local minima.

The fact that in large domains of the parameter space, arbitrarily close starting points can converge to different solutions, can be undesirable in practical design work. It is therefore worth investigating how the behavior of optimization can be stabilized. The irregular character of the paths shown in Fig. 11 and the similarity between the basins obtained with OPTSYS [Fig. 4(b)], and the basins obtained with CODE V and ZEMAX [Figs. 8(a), (d) and (e)], indicate that the damping, which is used in these programs, is too low for stability in this example. The results shown in Fig. 4 suggest that in order to make the boundaries more regular, a higher damping factor in the optimization algorithm is needed. However, in the versions of CODE V and ZEMAX available at the time of this writing, the user does not have the possibility to influence the damping factor, which is computed automatically.

To prevent the optimization algorithm from jumping unpredictably, we have implemented an external damping factor in the macro language of CODE V. Our external damping factor does not change the direction of the optimization step, but it decreases the step size. The direction of the variable change at each iteration is still determined by the optimization algorithm. Our procedure is illustrated in Fig. 12(a). Note that this technique is not equivalent to a change of the automatic internal damping of the optimization algorithm, but it removes the instabilities in the optimization process. For a discussion of a typical way to implement internal damping, which is to be preferred to external damping, see for example Ref. [5]. Our macro language implementation of external damping is about 6 times faster than our differential equation macro,

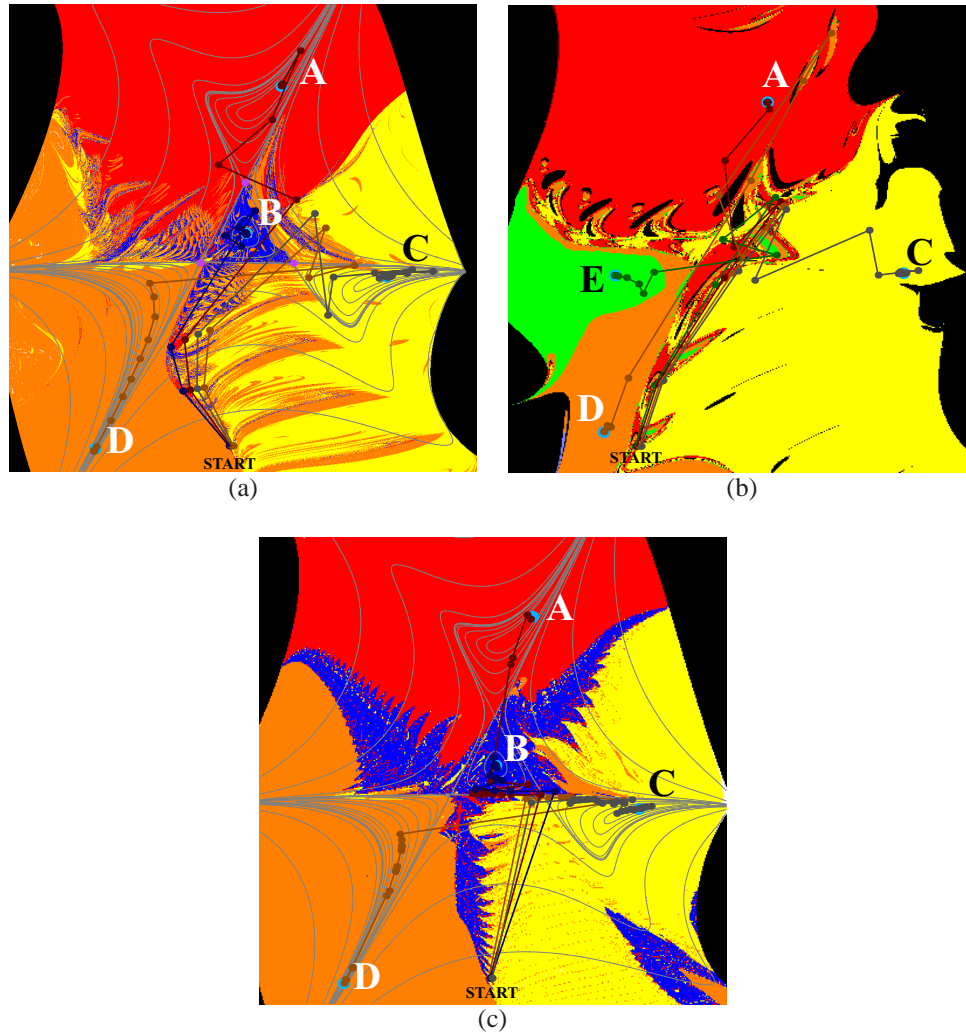


Fig. 11. Optimization paths (dark colored lines) in the variable space of the doublet, obtained with (a) default optimization of CODE V, (b) OPTSYS, (c) damped least-squares optimization of ZEMAX. Four starting points for optimization are chosen very close to each other at the position 'START', and the iterations are shown as dots. Note that the sequence of iterations always remains inside the same basin as for the starting point, independent of the basin shape.

but about 20 times slower than the standard optimization that is internally implemented in CODE V.

Figure 12(b) shows the basins of attraction in CODE V for the doublet with high external damping. In comparison with Fig. 8(a), the basin boundaries in Fig. 12(b) are regular and the basins are compact, which confirms our hypothesis that in Fig. 11(a) the chaotic transients are caused by a damping that is too low to produce a stable result. However, unlike internal damping, adjusting the external damping does not lead to basin shapes that are similar to the reference basins shown in Fig. 3(b).

When we compare Figs. 3(b), 8(a), and 12(b), we observe that the saddle points in the merit

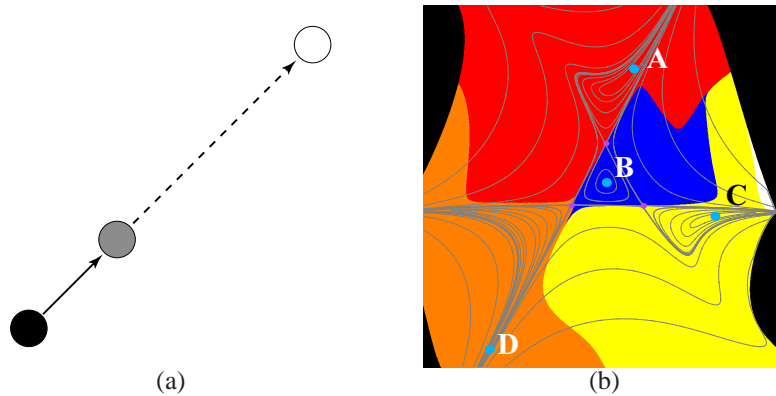


Fig. 12. (a) Illustration of the external damping procedure. The starting point is shown in black, and the result obtained after one optimization step with CODE V is shown in white. The result after external damping is shown in gray. The external damping factor damps the optimization in such a way that it does not change the direction, but only shortens the step size. (b) Basins of attraction for the doublet after applying our external damping to the default optimization method of CODE V.

function landscape are unaffected by the choice of algorithm or damping. Although the boundary shapes vary considerably when the optimization algorithm or the applied damping are altered, the saddle points always remain points on the basin boundaries.

5. Conclusion

A fundamental way to understand when optimization leads to a certain local minimum rather than to a neighboring one, is to look at the basin of attraction for that minimum. The examples discussed in this paper show that the basin structure and boundaries depend on the implementation details of the algorithm that is used and on numerical parameters, such as the damping factor. However, the saddle points in the merit function landscape are always on a basin boundary.

We show that, in certain situations, damped least-squares optimization methods can have an unpredictable behavior. In our examples, there are regions in the merit function landscape where starting points, which are very close to each other, lead to different local minima after optimization. This instability illustrates an extreme sensitivity to change of initial configurations. Computations of the capacity dimension then show that the basin boundaries have a fractal shape. When optimization is started in a fractal region, it first displays a chaotic transient path in the merit function landscape, before it finally converges to a local minimum.

By using damped least-squares methods with sufficient damping, we can obtain smooth basin boundaries even in optimization problems that have a natural propensity for fractal basins. Depending on the specific optimization problem and optimization method that is used, instabilities may be absent or present in different degrees. If these instabilities are present, then the main cause for them is to be found in the optimization algorithm.

The concept of damping factor was originally introduced with the purpose of limiting the step size during optimization, so that certain mathematical approximations in the least-square algorithm remain valid. Many optimization algorithms presently use much lower values of the damping factor. Although the exact details may vary from one program to another, the automatic choice of the damping factor is typically made such that the merit function decrease per optimization step is maximized. However, as we have shown, the gain in computation speed is

