Chaotic behavior in an algorithm to escape from poor local minima in lens design

Maarten van Turnhout and Florian 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: In lens design, damped least-squares methods are typically used to find the nearest local minimum to a starting configuration in the merit function landscape. In this paper, we explore the use of such a method for a purpose that goes beyond local optimization. The merit function barrier, which separates an unsatisfactory solution from a neighboring one that is better, can be overcome by using low damping and by allowing the merit function to temporarily increase. However, such an algorithm displays chaos, chaotic transients and other types of complex behavior. A successful escape of the iteration trajectory from a poor local minimum to a better one is associated with a crisis phenomenon that transforms a chaotic attractor into a chaotic saddle. The present analysis also enables a better understanding of peculiarities encountered with damped least-squares algorithms in conventional local optimization tasks.

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. G. W. Forbes and A. E. Jones, "Towards global optimization with adaptive simulated annealing," in *1990 Intl Lens Design Conf*, G. N. Lawrence, ed., vol. 1354 of *Proc. SPIE*, 144–153 (1991).
2. T. G. Kuper and T. I. Harris, "Global optimization for lens design - an emerging technology," in *Lens and optical system design*, H. Zuegge, ed., vol. 1780 of *Proc. SPIE*, 14–28 (1992).
3. M. Isshiki, H. Ono, K. Hiraga, J. Ishikawa, and S. Nakadate, "Lens design: Global optimization with Escape Function," Optical Review (Japan) **6**, 463–470 (1995).
4. K. E. Moore, "Algorithm for global optimization of optical systems based on genetic competition," in *Optical Design and Analysis Software*, R. C. Juergens, ed., vol. 3780 of *Proc. SPIE*, 40–47 (1999).
5. L. W. Jones, S. H. Al-Sakran, and J. R. Koza, "Automated synthesis of both the topology and numerical parameters for seven patented optical lens systems using genetic programming," in *Current Developments in Lens Design and Optical Engineering VI*, P. Z. Mouroulis, W. J. Smith, and R. B. Johnson, eds., vol. 5874 of *Proc. SPIE*, 587403 (2005).
6. J. P. McGuire, Jr., "Designing easily manufactured lenses using a global method," in *International Optical Design Conference 2006*, G. G. Gregory, J. M. Howard, and R. J. Koshel, eds., vol. 6342 of *Proc. SPIE*, 63420O (2006).
7. J. R. Rogers, "Using global synthesis to find tolerance-insensitive design forms," in *International Optical Design Conference 2006*, G. G. Gregory, J. M. Howard, and R. J. Koshel, eds., vol. 6342 of *Proc. SPIE*, 63420M (2006).
8. O. Marinescu and F. Bociort, "Network search method in the design of EUV lithographic objectives," Appl. Opt. **46**, 8385–8393 (2007).
9. D. C. Sinclair, "Optical design software," in *Handbook of Optics, Fundamentals, Techniques, and Design*, M. Bass, E. W. Van Stryland, D. R. Williams, and Wolfe W. L., eds., vol. 1, 2nd ed., 34.1–34.26 (McGraw-Hill, New York, 1995).
10. M. Laikin, *The method of lens design*, Lens design, 4th ed. (CRC Press, Boca Raton, FL, 1996).

11. A. Serebriakov, F. Bociort, and J. Braat, "Finding new local minima by switching merit functions in optical system optimization," Opt. Eng. **44**, 100,501 (2005).
12. H. Gross, H. Zügge, M. Peschka, and F. Blechinger, *Principles of optimization*, in *Handbook of Optical Systems*, vol. 3, 291–370 (Wiley-VCH, Weinheim, 2007).
13. D. Shafer, "How to optimize complex lens designs," Laser Focus World **29**, 135–138 (1993).
14. D. Shafer, "Global optimization in optical design," Computers in Physics **8**, 188–195 (1994).
15. M. van Turnhout and F. Bociort, "Instabilities and fractal basins of attraction in optical system optimization," Opt. Express **17**, 314–328 (2009). URL http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-17-1-314.
16. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
17. E. Ott, *Chaos in dynamical systems*, 2nd ed. (Cambridge University Press, Cambridge, 2002).
18. H. E. Nusse and J. A. Yorke, "Basins of attraction," Science **271**, 1376–1380 (1996).
19. S. N. Rasband, *Chaotic dynamics of nonlinear systems* (Wiley, New York, 1990).
20. I. Castillo and E. R. Barnes, "Chaotic behavior of the affine scaling algorithm for linear programming," SIAM J. on Optimization **11**, 781–795 (2000).
21. B. Davies, "Period Doubling," in *Encyclopedia of Nonlinear Science*, A. Scott, ed. (Routledge, New York, 2004).
22. C. Grebogi, E. Ott, and J. A. Yorke, "Chaos, strange attractors, and fractal basin boundaries in non-linear dynamics," Science **238**, 632–638 (1987).
23. Y.-C. Lai and C. Grebogi, "Converting transient chaos into sustained chaos by feedback control," Phys. Rev. E **49**, 1094–1098 (1994).

1. Introduction

Optical system designers are usually confronted with many local minima in the merit function landscape of their design problem. A major challenge is to find good solutions among these minima. Increasingly elaborate and powerful global optimization methods [1, 2, 3, 4, 5, 6, 7, 8] can provide a remedy when local optimization produces an unsatisfactory solution. However, simpler empirical techniques can also be effective for escaping from poor local minima, especially when the merit function barrier that must be overcome is low. Such techniques include reoptimization after making small changes to lens parameters, changing weights in the merit function, switching merit functions during local optimization, and modifying the conditions under which local optimization algorithms operate, for example by modifying the damping factor or even changing the method of local optimization [9, 10, 11].

In the past decades, various versions of the damped least-squares algorithm have been successfully used in optical system design [9, 12]. If used in the conventional way, damped least-squares algorithms search only locally for a minimum and are therefore computationally much less expensive than global optimization methods.

Shafer has shown that controlling the damping factor of a lens design program can help the designer to find alternative system configurations [13, 14]. Similarly, in this paper, we use low damping in a damped least-squares algorithm as an empirical technique to escape from poor local minima, and we temporarily allow configurations with higher merit function than for the initial configuration. For such empirical techniques, simplicity is their main asset and raison d'être, given the fact that they are not as powerful as global optimization methods. We have chosen this particular strategy because it was very simple to implement it in an existing code.

Since we allow the merit function to increase during the search process, the algorithm does not even always converge and more complicated scenarios including chaos are also observed in the asymptotic regime. A simple example will show that when this strategy does lead to a better configuration on the other side of a merit function barrier, this does not happen in a straightforward way. The new configuration will be obtained after a complex process including a chaotic transient. Because of the novelty within the framework of optical design methodology of the mechanism we have observed, the emphasis in this paper will be on the anatomy of this complex mechanism, rather than on the practical aspects of such a method.

The change of the damped least-squares algorithm discussed here is also useful for a different, but not less important purpose. We have recently found unexpected instabilities in the

behavior of conventional damped least-squares local optimization, even in computer programs that are widely used at present [15]. A better understanding of these instabilities is necessary in order to assess whether, or in what degree, these instabilities affect design productivity. We will show that the present algorithm change facilitates the study of certain features of these instabilities.

Our modified implementation of the Levenberg-Marquardt damped least-squares algorithm is shortly described in Sec. 2. In Secs. 3 and 4, we describe the behavior of this algorithm when different values of the damping factor are used. Section 3 shows how a chaotic attractor is formed, and Sec. 4 shows that a successful escape from a poor local minimum to a better one is associated with destruction of the chaotic attractor via what is called a 'crisis' in chaos theory. In Sec. 5, we use the obtained results for a better understanding of the instabilities observed in the behavior of conventional damped least-squares local optimization when used with low damping.

2. Algorithm with adaptive damping for searching beyond local minima

In this work, we use the program OPTSYS written by Joseph Braat, in which the Levenberg-Marquardt damped least-squares method is implemented as described in detail in Ref. [16]. We first describe briefly the conventional use of this code for local optimization. During optimization, the damped least-squares algorithm dynamically controls the damping factor. The algorithm uses two loops to step through the merit function landscape. At every optimization step, an outer iteration loop calculates the derivative matrix $\mathbf{J}$ of the optimization problem to find the search direction in which the value of the merit function decreases. This derivative matrix $\mathbf{J}$ is subjected to a singular value decomposition (SVD) with complete orthogonalization of both the operand space (dimension m) and the variable space (dimension n).

An inner loop uses the known matrices to optimize the damping factor λ. Successive linear damped solutions are then obtained by solving the system

$$(\mathbf{J} - \lambda_k \mathbf{I}) \, \mathbf{x}_k = -\mathbf{f}, \tag{1}$$

for different values of λ_k, where the index k denotes the cycle number of the inner loop. $\mathbf{I}$ is the identity matrix, zero padded where needed given the dimensions (m, n) of the system and $\mathbf{f}$ is the vector of operand values to be reduced to zero or certain target values. The vector of operand values is extended when necessary with the weighted values of constraint violations. For more details, see Ref. [16].

Because every optical design program has his unique way of choosing the damping method, there is a large diversity in the implementation details of local optimization [9]. In OPTSYS, the choice of the damping factor λ_k is done as follows. For the k-th iteration in the inner loop, λ_k is given by:

$$\lambda_k = p \, S_1 \, 10^{-(k-1)/a}, \tag{2}$$

where S_1 is the largest singular value delivered by the SVD and p and a are values to be provided by the user. In this work, we have chosen a fixed value $a = 10$, but the value of p has been varied. In the following two sections, p will be the control parameter that determines whether the algorithm converges or whether its behavior is more complex. During optimization, the algorithm monitors the decrease in the squared value $|\mathbf{f}_k|^2$ of the merit function as a function of the counter k and stops the iteration of the inner loop when the merit function starts increasing again.

To keep the calculation time short, the algorithm does not carry out an interpolation to find the solution $\mathbf{x}$ corresponding to the exact local minimum of the merit function. The quasi-minimal linear solution $\mathbf{x}_k$ is used to update the system variables. The new working point is then used to obtain the derivative matrix that serves for the next (outer) optimization cycle.

A slight change in the conventional form of this algorithm is sufficient to overcome, for low damping, the merit function barrier that separates a poor local minimum from a better one, as will be seen in Sec. 4. We iterate the inner loop at least two times and we allow the merit function to increase, but only in the first step of the inner loop. After the first two iterations, the process is stopped when $|\mathbf{f}_{k+1}|^2 > |\mathbf{f}_k|^2$.

3. Period doubling route to chaos

When we want to escape from a poor local minimum and achieve convergence towards a different, hopefully better one, we first have to destabilize the convergence towards the original local minimum. An intermediate stage, which is discussed in this section, is the formation of a chaotic attractor. Depending on the starting configuration, the stages between convergence and chaos can be different. For simplicity, in this section we illustrate the well-known period doubling route to chaos.

We examine the behavior of the algorithm described in the previous section for a simple two-dimensional monochromatic doublet optimization problem (f number 3, field of 3 degrees), with the curvature of the second and third surfaces (c_2 and c_3, respectively) as variables. For sufficiently large damping, optimizing the two variable curvatures results in five different local minima (A, B, C, D, and E), which are shown in Fig. 1. More details about this optimization problem are given in a recent open access paper [15]. The observed behavior depends on three factors: the choice of the initial configuration, the value of the damping parameter p, and the number of iterations that have been performed. Therefore, in this paper, we present three sorts of numerical results, in which one of these factors is varied.
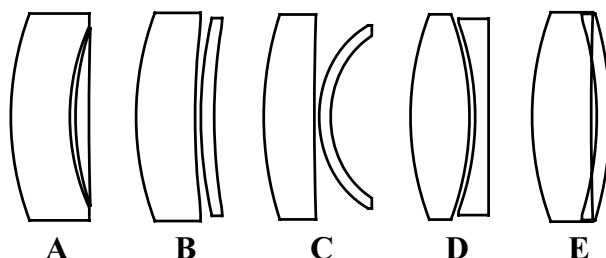

Fig. 1. Five local minima for a simple two-dimensional monochromatic doublet optimization problem. The control of edge thickness violation was disabled in order to eliminate the constraints from the list of possible causes of the peculiar behavior described in what follows.

In this section, we are especially interested in the result of the iteration process in the asymptotic regime (i.e. after a large number of outer iterations). When p is sufficiently large, the algorithm is convergent, as expected. For our example, in order to show which configuration leads to a certain local minimum, we compute the basin of attraction for that local minimum. (The set of all starting configurations that are attracted to a local minimum is the basin of attraction for that minimum [15, 17, 18].)

Figure 2 shows the basins of attraction for $p = 0.002$. In this and in the following figures of the same kind, curvatures c_2 and c_3 are plotted along the vertical and horizontal axis, respectively. The five local minima (see Fig. 1) are indicated by blue dots and the letters A, B, C, D, and E. Compared to the other minima, minimum B has a high merit function value.

The basins of attraction are computed by using a grid of 101×101 equally spaced starting points in the two-dimensional variable space, and at each grid point, we start to optimize the corresponding configuration, until it arrives in one of the local minima. To each of the five local
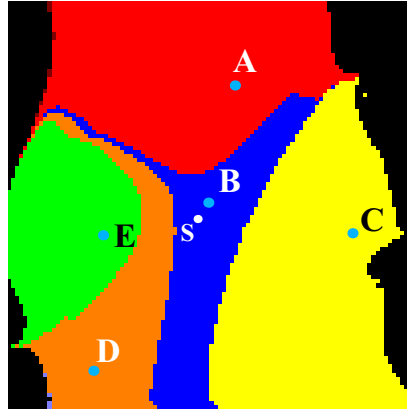
Fig. 2. Basins of attraction for the five local minima in Fig. 1, obtained with damping parameter $p = 0.002$ on a grid of 101×101 points. Each grid point is iterated 999 times. The starting point 'S' will be used in Sec. 4.

minima, we associate a different color: red, blue, yellow, orange, and green to local minima A, B, C, D, and E, respectively. Depending on which local minimum we obtain, we color the starting point with the corresponding color for that minimum. 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. If ray failure occurs during optimization, we stop the optimization process and color the corresponding starting point black as well. Note that in Fig. 2, the basins for all five minima have smooth boundaries.

When we decrease p, we observe types of behavior that are much more complex than the convergent behavior towards the minima shown in Fig. 2. In the asymptotic regime, the limit of the sequence of iterations is then not necessarily a point, but can be a set of points. For example, for a certain starting configuration in the former basin of local minimum C, below a critical value of $p \approx 0.000138$, the algorithm does not converge to local minimum C anymore. Figure 3 shows how the algorithm oscillates between different points in the variable space when we decrease the value for p. For each value of p, we computed 999 iterations. The values of curvature c_3 of the last 899 iterations are plotted superimposed as function of p (p decreases along the horizontal axis). Since we are interested in the asymptotic behavior, the first 100 iterations are discarded in the figure in order to avoid transitory features.

At $p_1 \approx 0.000138$, we observe a so-called pitchfork bifurcation [19]. Minimum C, which was stable for higher values of p, becomes unstable, but a stable pair of points (both of them distinct from C) appears. For p between 0.000138 and 0.000095, the algorithm alternates in the asymptotic regime between the upper and lower branch resulting from the bifurcation, which correspond to two points with different values of the merit function, without ever converging to any of them. This means that we have a periodic attractor consisting of two points. (An attractor is the set of points in the two-dimensional variable space towards which starting configurations in the corresponding basin of attraction approach asymptotically in the course of the outer iteration process.)

In the asymptotic regime, the iteration process has now a period of two. By starting the algorithm at one of the attracting points, each second iteration produces the starting point as result: after one (outer) iteration, we obtain the other point of the attractor; after the next iteration, the system is moved back to the original point, and so on. (In Fig. 3, the points on the vertical lines
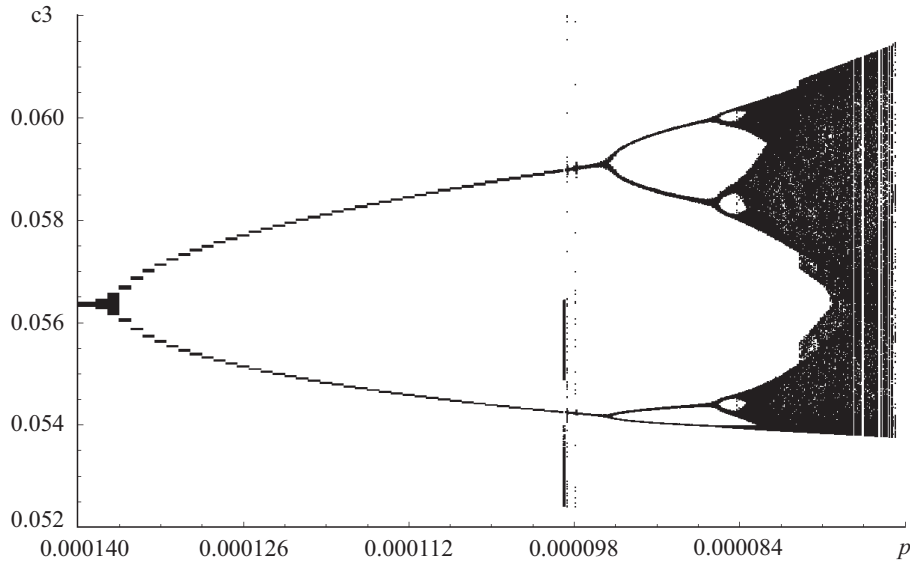
Fig. 3. Period doubling route to chaos. Curvature c_3 of the points in the asymptotic regime for a starting configuration in the basin of doublet local minimum C (see Fig. 2) is shown as function of the damping parameter p.

close to $p = 0.000100$ can be discarded.)

When p is decreased further, the two points bifurcate again into four points at $p_2 \approx 0.000095$ (period of four), then at $p_3 \approx 0.000086$ into eight points (period of eight). In the latter case, only seven out of eight distinct points can be easily distinguished in Fig. 3, because two points in the attracting set have approximately the same value for c_3. Examining the values of curvature c_2 shows that these points are distinct. Note that the values of p where the bifurcations take place, become closer and closer to each other, until p reaches a critical value at which a chaotic set of attracting points appears. This behavior is known as the period doubling route to chaos, the most common of several routes to chaos for a nonlinear dynamical system [19].

The pitchfork bifurcations in Fig. 3 are similar to the ones found in the logistic map and in experiments in many areas, including hydrodynamics, electronics and laser physics. Chaotic behavior is also observed in algorithms for linear programming [20]. In Fig. 3, the ratio of the first two bifurcation intervals

$$\delta = \frac{p_2 - p_1}{p_3 - p_2} \approx 4.78 \tag{3}$$

is close to the value of the universal Feigenbaum constant ($\delta = 4.6692...$) encountered for functions that approach chaos via period doubling [21]. After a range of chaotic behavior, for values of p smaller than approximately 0.000071, the algorithm encounters ray failures and the iterative process is stopped.

The period doubling route to chaos is also illustrated in Fig. 4. Figures 4(a)–(c) show periodic attractors consisting of two, four and eight points, respectively. The figures on the left-hand side show the iteration trajectories in the two-dimensional variable space. Curvatures c_2 and c_3 are plotted along the vertical and horizontal axis, respectively. The iteration points shown in gray are transients, and local minimum C is shown as a large gray point.

The figures on the right-hand side show the values of c_2 as function of the number of iterations, including the transients (gray points). After the transients, Figs. 4(a)-(c) show horizontal lines, and each line corresponds to the c_2 value for a point in the periodic attractor. As
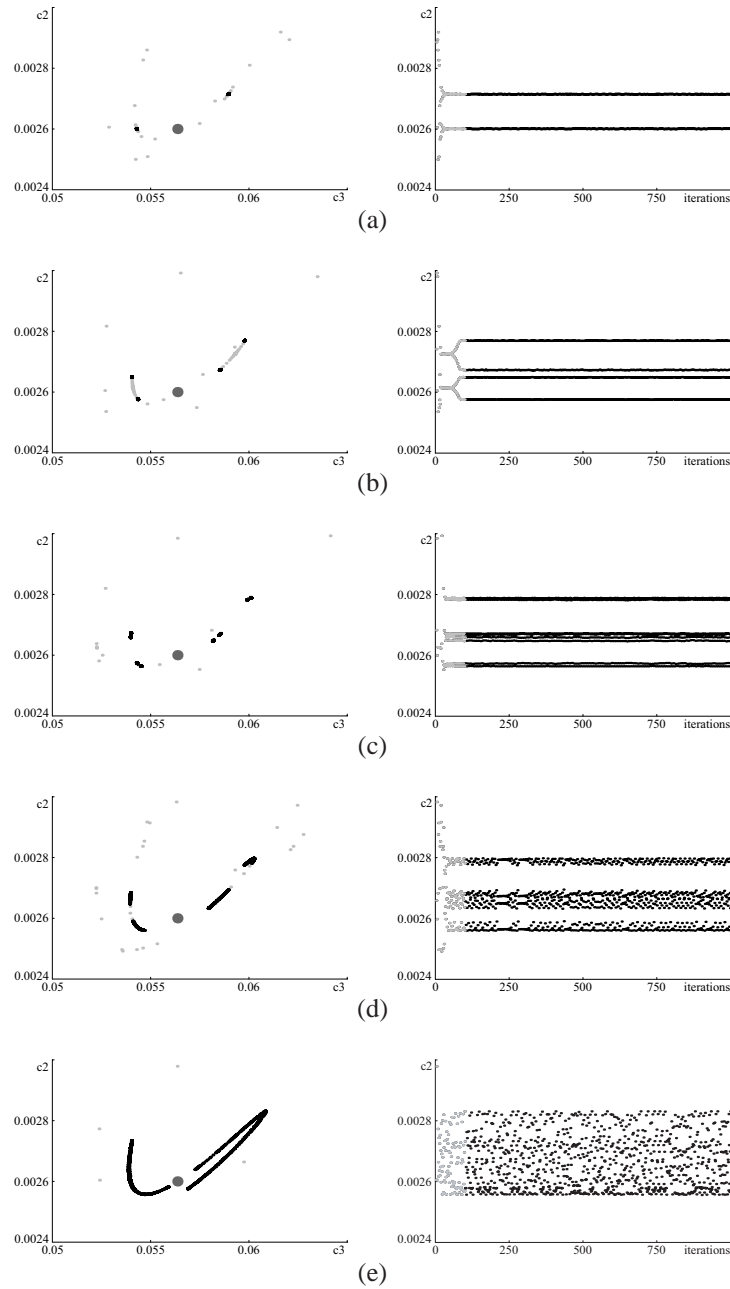
Fig. 4. Figures on the left: Iteration trajectories in the two-dimensional variable space (c_3, c_2) of a monochromatic doublet for five different values of the damping parameter p. The starting configuration is in the former basin of local minimum C (see Fig. 2). The large gray point corresponds to local minimum C (obtained with sufficient damping), and the iterations shown in gray are considered as initial transients. Figures on the right: The evolution of curvature c_2 as function of the number of iterations. (a) $p = 0.0001$, (b) $p = 0.00009$, (c) $p = 0.0000854$, (d) $p = 0.0000833$, (e) $p = 0.0000777$.
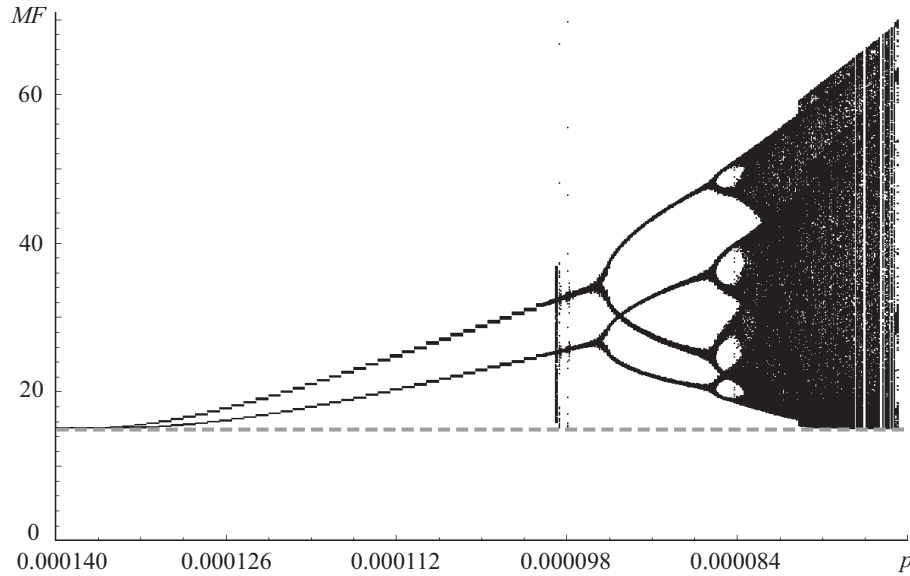
Fig. 5. Merit function (*MF*) values (in relative units) for the points shown in Fig. 3.

mentioned above, attracting points that have equal values of c_2 cannot be distinguished. For example, two horizontal lines overlap each other in Fig. 4(c). Figure 4(d) shows a remarkably complex pattern, corresponding to the transition region between order (period doubling) and chaos. In Fig. 4(e), the set of attracting points forms a chaotic attractor, which is similar to those studied in nonlinear dynamics [22].

It is important to note that in the examples shown in Figs. 3 and 4 the points of a certain periodic or chaotic attractor do not have the same value of the merit function. Figure 5 shows that the points of periodic or chaotic attractors obtained with low damping have a merit function value that is higher than or equal to the merit function value of the original minimum that was destabilized (dashed gray line). An algorithm change that allows the merit function to increase is therefore essential for non-convergent behavior.

By not allowing the merit function to increase during the process, in the conventional local optimization algorithms the set of attracting points must have the same merit function value in the asymptotic regime. Therefore, an attractor consisting of several points with different values of the merit function cannot exist. In the asymptotic regime, we have then a point attractor, which means that the final result is convergence. However, as discussed in Sec. 5, situations exist where the iterations of conventional local optimization algorithms based on damped least-squares methods first display an unpredictable chaotic path in the merit function landscape, before they finally converge to a local minimum. This type of behavior is known as a chaotic transient [17]. In the next section, we give another example of a chaotic transient.

4. Escaping from a poor local minimum

In this section, we show that the algorithm described in Sec. 2 may improve the result of local optimization that would otherwise converge to a poorer solution (local minimum B). When we use sufficiently low damping, the algorithm overcomes the merit function barrier between local minima B and A, and finally converges to the better minimum A. After a chaotic attractor is formed, further decrease of damping destroys the chaotic attractor and the iteration trajec-
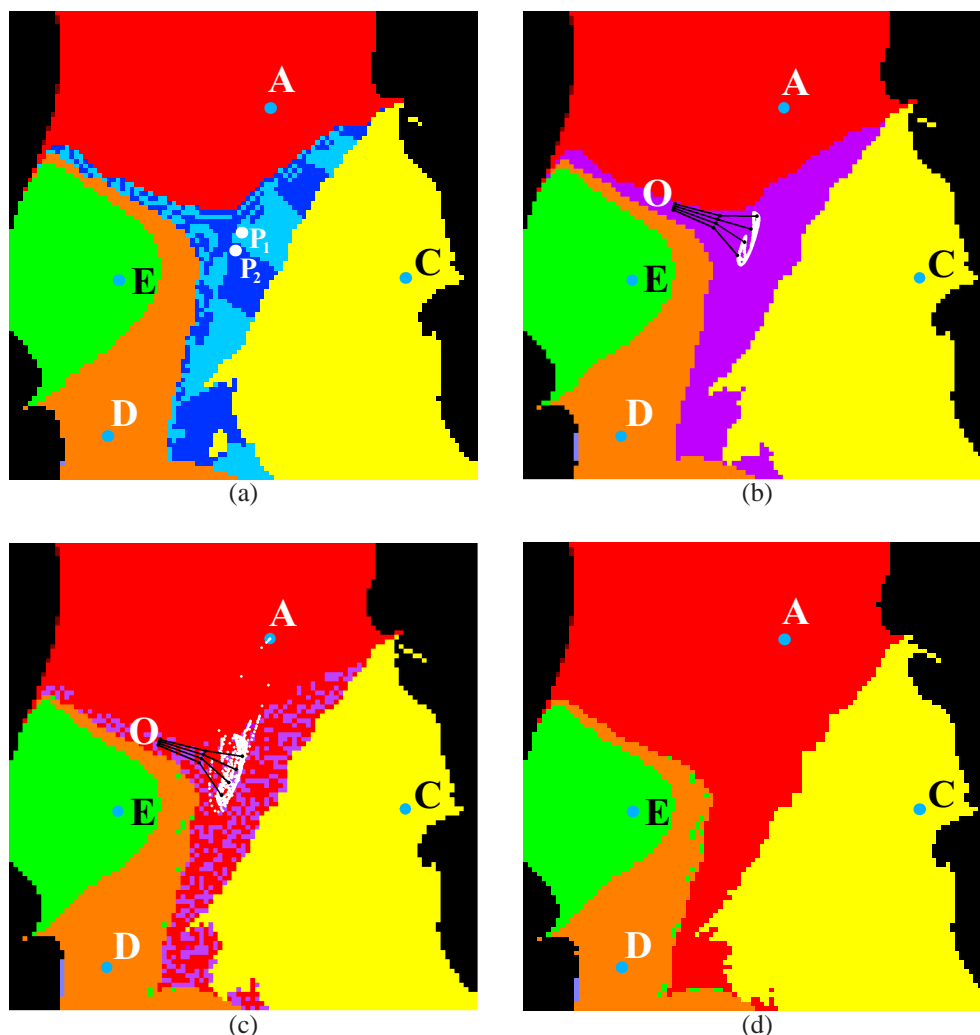
Fig. 6. Basins of attraction obtained for different values of the damping parameter p on a grid of 101×101 points. Each grid point is iterated 999 times. (a) $p = 0.0009$, (b) $p = 0.00075$, (c) $p = 0.000679$, and (d) $p = 0.0006$. The black lines starting close to point O in Figs. 6(b) and (c) will be explained in Sec. 5.

tory moves off to a different region of the variable space, where it converges to minimum A. However, before the iterations converge to minimum A, first a chaotic transient is observed.

Figure 6 shows the basins of attraction for the monochromatic doublet for four low values of p. We used a grid of 101×101 points, where each grid point is iterated 999 times. Note that the basins of minima A, C, D, and E remain compact. However, the basin of minimum B, which has the largest value of the merit function, completely disappears for $p = 0.0006$. First, for a variation domain of p including the value $p = 0.0009$, the algorithm approaches a periodic attractor consisting of two points [the white points P_1 and P_2 in Fig. 6(a)], which are close to the former point B. After a sufficiently large number of iterations n, the succession of each second iteration converges to P_1 (the regions that lead to this point after n large and odd are shown in dark blue) or to P_2 (light blue).

For even lower values of p, the number of attracting points increases until a chaotic attractor is formed [white points in Fig. 6(b)]. The purple region shows the set of starting configurations that are attracted to the chaotic attractor. When the damping is lowered further, the process iterates over the merit function barrier and converges to local minimum A [Fig. 6(c)–(d)]. In chaos research, one commonly encounters situations called 'crises' where, after the control parameter passes through a critical value, the chaotic attractor becomes unstable [22]. The same phenomenon is observed here.

When we choose as starting point for instance the white point indicated with 'S' in Fig. 2, or the point 'O' in Fig. 6(c) (in the set of four neighboring black points, O is the second one from below), the algorithm first displays a chaotic transient. Then, it oscillates for many iterations in a domain of the variable space that looks very similar to a set of attracting points, but which finally turns out to be unstable. In chaos research, such a set of points resulting from a chaotic attractor that became unstable after a 'crisis' is called a chaotic saddle [23]. Starting configurations for which the algorithm converges to minimum A are shown in red. The white points in Figs. 6(c) show the set of iteration points of the chaotic transient and of the subsequent escape to minimum A, when we choose point O as starting point (the first two iterations started from O are shown in black and connected with black lines). For the starting configurations colored in purple, the total number of 999 iterations was not sufficient for $p = 0.000679$ to escape from the chaotic behavior. When we increase the number of iterations, the configurations corresponding the the purple points also reach minimum A. When p becomes sufficiently low ($p = 0.0006$), all starting configurations in the former basin of minimum B converge within 999 iterations to minimum A [Fig. 6(d)].

The transition to chaos, the chaotic-like behavior, and the chaotic transients are also illustrated in Figs. 7 and 8. For the starting point S (see Fig. 2), Fig. 7 shows the transition from convergence (to minimum B) to chaos and then to convergence to minimum A after the 'crisis'. The first 300 iterations are discarded and the values of c_2 of the last 699 iteration points are plotted as a function of p. The horizontal lines on the left- and right-hand side of Fig. 7 correspond to the c_2 values of minima B and A, respectively.

Figure 8 shows the iteration trajectories for a set of values of p in the (c_3, c_2)-space (left), and the values of c_2 for the iteration points as function of the number of iterations (right). The iterations shown in gray are considered to be initial transients, and local minimum B is shown as a large gray point.

In comparison to Fig. 3, the route to chaos in Fig. 7 is a more unusual one. First, we observe a bifurcation, which is then followed by trifurcation. Note that the figure shown on the right in Fig. 8(a), where two stable attracting points can be observed, also seems to suggest the presence of six unstable fixed points. When p decreases, these six points become stable and the algorithm continues with oscillations between these six points [Fig. 8(b)].

In Fig. 8(c), the complex pattern of iterations shows that we are near the transition to chaos. For slightly smaller values of p, the iteration points form a chaotic-like set in the variable space [Fig. 8(d)]. After the 'crisis', this set of points becomes an unstable chaotic saddle. The algorithm escapes from it and moves to a different region of the variable space [Fig. 8(e)]. The convergence to local minimum A is illustrated by the horizontal line at the top right of Fig. 8(e). (Certain details are different when we choose another configuration in the former basin of minimum B as starting point.)

Note in Figs. 7 and 8(a)–(d) how the size of the stable periodic or chaotic attractors increases when the damping is reduced. For the chaotic attractor, the basin of attraction is the purple region shown in Fig. 6(b). As can be seen in Fig. 6(c), the 'crisis' occurs when the growing chaotic attractor reaches the chaotic basin boundaries. After this so-called 'boundary crisis' [17, 23], the white iteration points leave the mixed red-purple region [essentially the basin of the
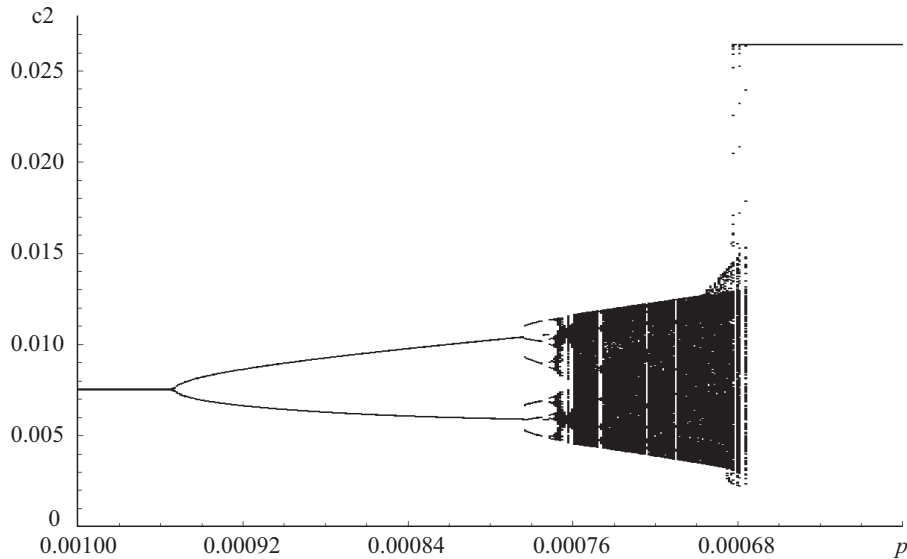
Fig. 7. Attracting points for the starting configuration S in the basin of local minimum B (see Fig. 2). The first 300 iterations have been discarded. Curvature c_2 of the iterations between 301 and 999 has been plotted as function of p.

chaotic attractor shown in Fig. 6(b)] and enter the compact red region (the basin of minimum A). The final result is then convergence to minimum A.

As in the case of Fig. 6, Fig. 9 shows that before the 'crisis' all points of periodic or chaotic attractors have a merit function which is higher than or equal to that of minimum B (dashed gray line). However, after the 'crisis' and after escaping from the chaotic saddle, the merit function decreases significantly and becomes that of minimum A (horizontal black line in the lower right part of the figure). A behavior similar to the one described above can also be observed when the algorithm escapes from minimum E to the somewhat lower minimum D. However, we then encounter ray failure for many values of p after the 'crisis'.

5. Instabilities in conventional damped least-squares optimization

In a recent open access paper [15], we have shown that optical design problems exist where conventional local optimization, as it is implemented in commercial optical design programs, is unstable if low damping is used. Then, starting points close to each other converge to different local minima. As we have shown there, if present, instabilities decrease the degree of predictability of the design process, and therefore a better understanding of the mechanisms by which such instabilities can arise is desirable.

While the detailed mechanisms of these instabilities require further study, below we show that certain features mentioned in Ref. [15] can be understood by examining Figs. 6(b)–(c). (If necessary, for observing details better, these figures can be enlarged on-screen in the electronic version of this article.) These figures show the second and third iteration for a set of four starting points close to each other (the black lines). The four trajectories are attracted by the chaotic attractor in Fig. 6(b) and initially also by the chaotic saddle in Fig. 6(c). While for the settings of Fig. 6(b) the attractor is stable and all four trajectories remain there indefinitely, in Fig. 6(c), the previous attractor that is now a chaotic saddle first attracts, then repels the trajectories, which finally arrive at minimum A. We note that, until the four trajectories reach the chaotic
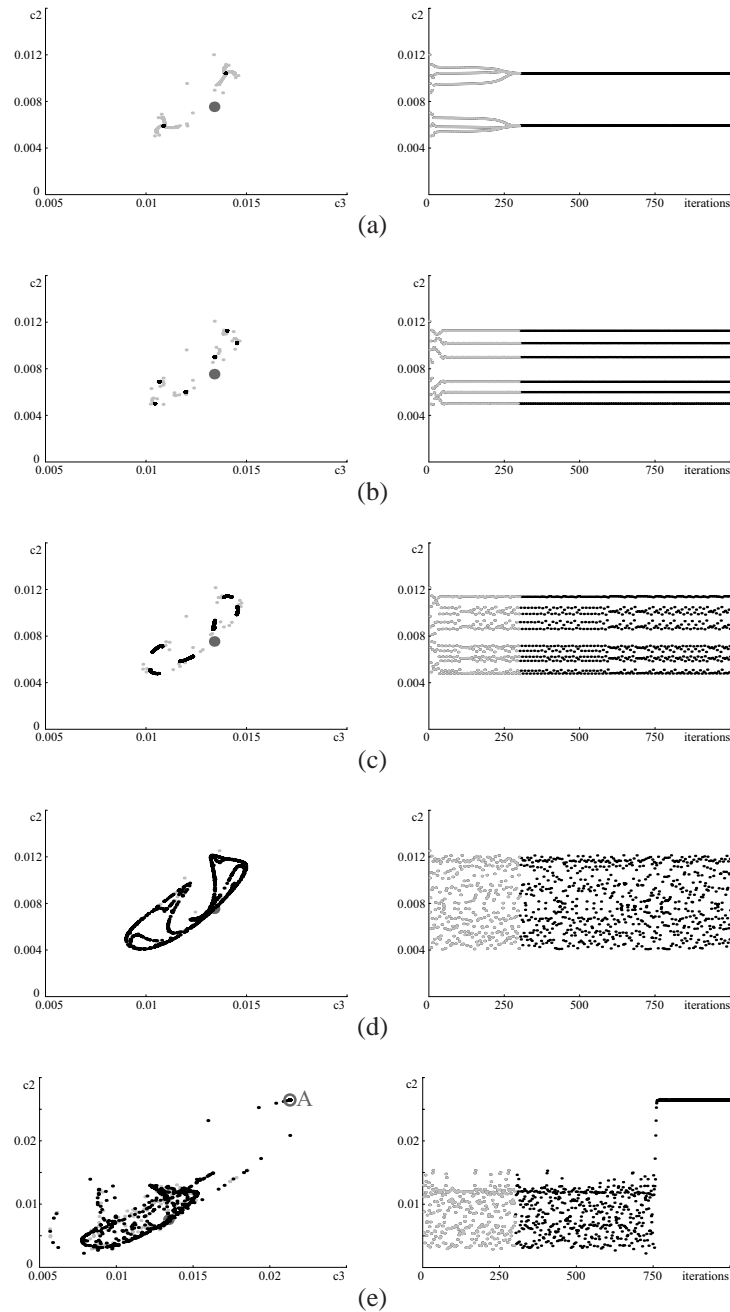
Fig. 8. Figures on the left: Iteration trajectories in the two-dimensional variable space (c_3, c_2) of a monochromatic doublet for five different values of the damping parameter p. The starting configuration is point S in the former basin of local minimum B (see Fig. 2). The large gray point corresponds to local minimum B (obtained with sufficient damping), and the iterations shown in gray are considered as initial transients. Figures on the right: The evolution of curvature c_2 as function of the number of iterations. (a) $p = 0.000784$, (b) $p = 0.000776$, (c) $p = 0.000768$, (d) $p = 0.000730$, (e) $p = 0.000679$.
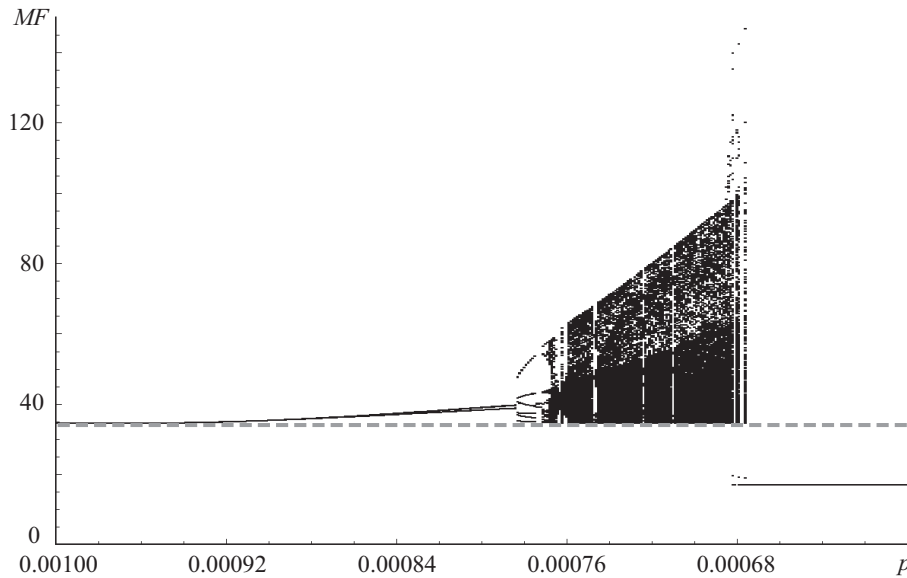
Fig. 9. Merit function (*MF*) values (in relative units) for the points shown in Fig. 7.

attractor in Fig. 6(b) and the chaotic saddle in Fig. 6(c), the distance between them increases significantly after each iteration.

A similar behavior can be observed in Figs. 11(a) and 11(c) of Ref. [15] and in the movie in Fig. 10 (Media 1) where essentially the same doublet optimization problem as the one discussed in this paper is analyzed with two different widespread commercial optical design programs (see Ref. [15] for full details). The movie shows the history of a set of ten starting points that are very close to each other. For each starting point, the iteration trajectories are first attracted, and then repelled by domains close to the red and purple lines in the movie.

The distance between several starting points close to each other increases significantly, until the trajectories reach a domain of the variable space that is close to a certain equimagnitude contour line (i.e. a curve along which the merit function is constant) that first attracts, then repels the trajectories. For instance, a large number of iteration points in the middle of Fig. 11(c) of Ref. [15] indicates the presence of a horizontal attracting-repelling line there.

The extreme sensitivity to initial conditions occurs because, when they have reached the attracting-repelling lines, these trajectories are already far apart from each other and therefore, after the trajectories are repelled, they converge to different local minima [shown in blue in Ref. [15] and Fig. 10 (Media 1)], despite of the fact that they have their origin in almost the same point.

Since the domains close to the attracting-repelling lines play a key role in the mechanism of the instabilities, it is important to understand their nature. However, studying chaotic transients directly in commercial programs is difficult, because there the chaotic transients are short. Since it can be used to increase the duration of the chaotic transients, the modification of the damped least-squares algorithm described in this paper in helpful for an analogy. As shown in chaos research, there is a fundamental relationship between the presence of chaotic transients and the existence of a chaotic saddle [23]. This fact, and the fact that the domains close to the attracting-repelling lines in the movie act in a way that is similar to what is proven to be a chaotic saddle in Fig. 6(c), suggest that these domains contain a chaotic saddle (or parts of it) as well.
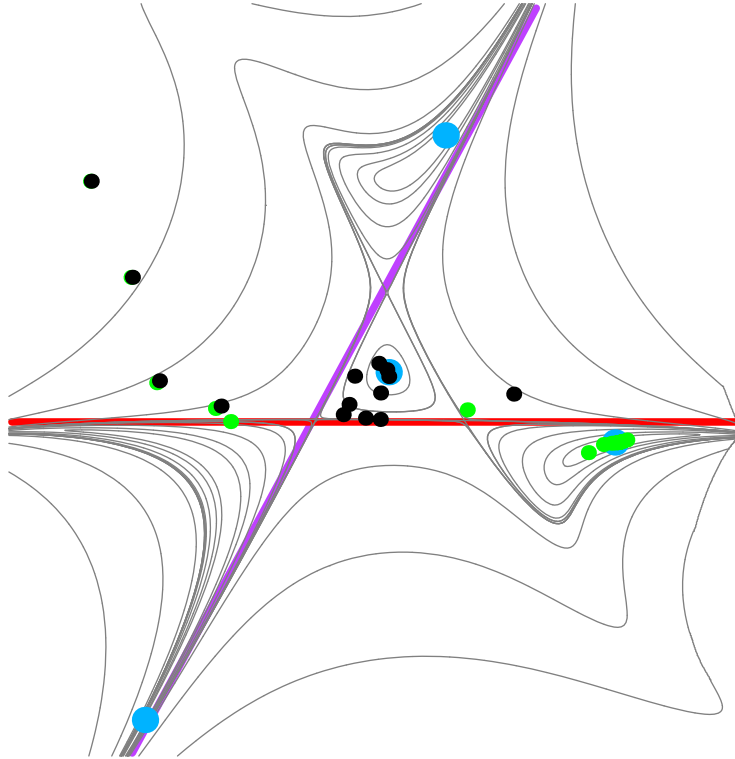
Fig. 10. ([Media 1]). Iteration history obtained with CODE V data of a set of starting points that are very close to each other. The merit function equimagnitude contours (i.e. the contours along which the merit function remains constant) are shown in gray, and the history of earlier trajectories is shown in green.

6. Conclusion

Our simple example shows that a merit function barrier in the optical design landscape can be overcome by using a damped least-squares method with reduced damping, in which the merit function is allowed to increase. Like most other empirical strategies of this kind, success is not guaranteed (ray failure sometimes occurs). However, when the present method is successful, the complex mechanism we have observed is one that is frequently encountered in chaos research. Because such mechanisms depend not on model details but rather on some general properties of the model, we believe that these results are relevant not only for the present algorithm with its specific implementation choices, but for other possible algorithms of the same kind as well.

To escape towards a better local minimum, it is necessary to destabilize first the convergence of the algorithm towards the original poor minimum, and we have seen that this can be achieved by lowering the damping. However, before convergence to a different minimum occurs, decreasing the damping leads to the successive stabilization and then destabilization of a sequence of periodic and chaotic attractors and to the growth of the size of the attractor. The distance between the attractor and its basin boundary decreases, and when the attractor touches its basin boundary, the attractor is destroyed by a boundary crisis.

For damping values that are lower than the critical crisis value, after a chaotic transient, the final result is convergence to a minimum which is distinct from the original one. The sequence of iterations is first attracted, then repelled by the chaotic saddle that results from the chaotic

attractor after the crisis.

The chaotic transients reported earlier [15] suggest that such complex behavior is more general, and also occurs with conventional damped least-squares algorithms, if they are designed to maximize speed. By not allowing the merit function to increase (significantly) during local optimization, the occurrence of periodic or chaotic sets of points with different values of the merit function, that might prevent convergence, is not possible, and conventional damped least-squares methods are convergent, as expected. However, in early stages of optimization, conventional low-damping algorithms can create chaotic transients, which could add an extra element of unpredictability in the design process. Because of their temporary character, chaotic transients are more difficult to be studied than chaos itself. The modification of the damped least-squares algorithm used here was also a mean to increase the duration of the chaotic transient, and even to stabilize the chaotic phase. Therefore, this modification is helpful for a better understanding of qualitative features of the unexpected behavior in widespread optical design programs.

The position in the variable space and the geometry of domains that act like chaotic saddles are important factors that determine whether instabilities in optimization are present or not. As we have seen in Fig. 6(c), the presence of chaotic transients and of a chaotic saddle do not necessarily lead to instabilities in optimization. The growing chaotic attractor in Fig. 6(b), that becomes a chaotic saddle in Fig. 6(c), exceeds its own basin boundary in only one region, and in the final phase of the chaotic transient the optimization converges to the minimum that has its basin on the other side of that region. Virtually all starting points in the basin of the former chaotic attractor converge finally to the same minimum and instabilities are not present. However, in the example shown in Sec. 5, starting points close to each other do converge to different minima because domains acting like chaotic saddles first attract neighboring trajectories (and at the same time enlarge the distances between them) and then repel them so that they leave the domain in different regions that are far away from each other.

Acknowledgments