© 2013 THOSS Media & **DE GRUYTER** DOI 10.1515/aot-2012-0057 — Adv. Opt. Techn. 2013; 2(1): 21–29

Tutorial

Masaki Isshiki*

Lens design: optimization with Global Explorer

Abstract: The optimization method damped least squares method (DLS) was almost completed late in the 1960s. DLS has been overwhelming in the local optimization technology. After that, various efforts were made to seek the global optimization. They came into the world after 1990 and the Global Explorer (GE) was one of them invented by the author to find plural solutions, each of which has the local minimum of the merit function. The robustness of the designed lens is also an important factor as well as the performance of the lens; both of these requirements are balanced in the process of optimization with GE2 (the second version of GE). An idea is also proposed to modify GE2 for aspherical lens systems. A design example is shown.

Keywords: aspheric system; optical design; optimization; tolerance.

OCIS codes: 220.3620; 220.0220.

***Corresponding author: Masaki Isshiki,** Isshiki Optics: 3-10-10 Terabun, Kamakura 247-0064, Japan, e-mail: isshiki@yk.netlaputa.ne.jp

1 Introduction

Designing a lens is somewhat like prospecting for a mineral vein in the crust of the earth, as depicted in Figure 1. Even if one finds a good vein, there is no way of knowing if a better one might exist nearby or far away. The situation is similar in lens design. If a designer finds a lens design using some optimization technique, the design thus found is only a local solution located near the starting design. The solution may be non-optimal if the starting point is inappropriate.

While the crust of the earth extends only in three spatial dimensions, the space over which the lens design solutions are distributed is a multidimensional parameter space.

The performance of an optical system can be evaluated by a single value called the 'merit function', which the designer seeks to minimize. In a conventional optimization, one aims to find a lens system with the smallest merit function near a staring design. Thus, the performance of the optimized lens depends heavily on the choice of starting point, which can be problematic. This problem was not solved until 1990, after which several methods of global optimization appeared [1–6]. One such global optimization method, called the Global Explorer [7–9] or GE can automatically find multiple local solutions from any starting design.

The major advantages of the GE method are:

1. It is fully automatic in searching for multiple solutions.
2. It is computationally efficient.
3. The designer's intentions can be easily implemented.
4. All the standard techniques of the damped least squares (DLS) method can be fully utilized.

The newest version of GE, GE2, optimizes both of the lens performance and the robustness in the optimization process. It is based on 'θ segmentation' and dramatically shortens the search time compared to GE.

Guidance in the design of lenses with aspheric surfaces is provided in order to cope with the special feature of such lenses.

2 Local optimization in lens design

2.1 Starting design

In designing a lens, one must first set up starting design specifications, such as the number of lens elements, their arrangement, the materials used, and so on. The choice of the starting design is critical to the performance of the solution reached. From a good starting point, the computer can easily find a solution automatically by the conventional optimization technique. However, there is no standard or systematic way to find a good starting design. Furthermore, the automatic optimization process cannot (1) divide one lens element into two or combine multiple elements into one; (2) separate cemented lens elements or cement together two or more elements; (3) change a spherical surface to an aspheric one or vice versa; (4) change a refractive element to diffractive one or vice versa; (5) select lens materials so as to take advantage of abnormal dispersion characteristics.

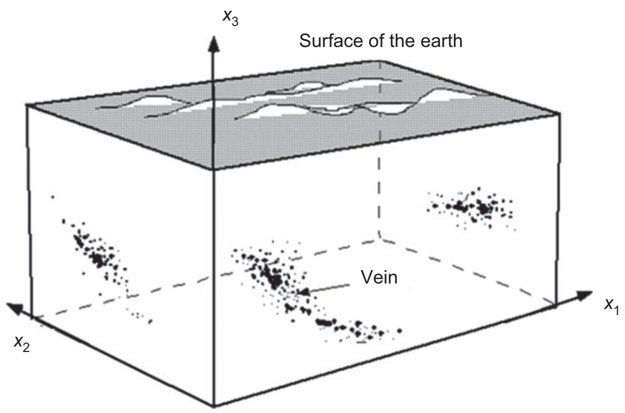

Figure 1 Vein searching in the mining industry.

2.2 Error functions

Error functions, designated as

$$f_i (i=1, 2,...,m), \tag{1}$$

characterize the deviation of an optical system from the ideal, for example, the deviation of the point at which an image-forming ray passes through the image plane from its ideal position. Design parameters are denoted by

$$x_j \ (j=1, 2,...,n). \tag{2}$$

These include surface curvatures, axial separations between surfaces, refractive indices of glasses, etc. Image errors are functions of x_j as indicated by

$$f_i(x_j) \ (i=1, 2,...,m). \tag{3}$$

Ray tracing is indispensable in arriving at error functions. Several approaches have been proposed to save time in image evaluation. One of these is described in the reference [13].

2.3 Damped least squared method

Designing a lens consists of finding a set of x_j that minimizes the error functions. The computer tries to minimize the value of the merit function ϕ defined by

$$\phi=\sum_{i=1}^{m} w_i\left(f_i\text{-}f_i^T\right)^2, \tag{4}$$

where f_i is an error function to be controlled, f_i^T is its target value (zero in most cases), and w_i is the weight for f_i. Hereafter, we assume that the weight and target value are included in each error function, so that the merit function can simply be written as

$$\phi=\sum_{i=1}^{m} f_i^2. \tag{5}$$

The condition for the minimized ϕ is given by

$$\frac{\partial\phi}{\partial x_j}=0 \quad (j=1, 2,\dots,n). \tag{6}$$

The design problem is to find those values for x_j that satisfy the above equations. This is not an easy task because ϕ is composed of error functions, f_i that are complex nonlinear functions of parameter x_j that can only be calculated numerically by ray tracing.

In order to solve this problem, f_i are linearly approximated in the vicinity of the starting point x_{j0} as

$$\bar{f}_i\left(x_j\right)=f_i\left(x_{10},\dots,x_{n0}\right)+\sum_{i=1}^{n} a_{ij}\left(x_j\text{-}x_{j0}\right), \tag{7}$$

where a_{ij} is

$$a_{ij}=\frac{\partial ff}{\partial x_j}\approx\frac{\Delta\ _i}{\Delta x_j} \text{ at } x_j=x_{j0}. \tag{8}$$

Using these, we have an approximated merit function

$$\bar{\phi}=\sum_{i=1}^{m} \bar{f}_i\left(x_j\right)^2, \tag{9}$$

which is a second-order polynomial in x_j. The values for x_j that minimize this approximated merit function can be obtained by solving the simultaneous linear equations

$$\frac{\partial\bar{\phi}}{\partial x_j}=0 \ (j=1, 2,\dots,n). \tag{10}$$

The solution thus obtained is an approximation that can be regarded as the new starting point for the next iteration of the above process. It is expected that such iteration will lead the design to converge to a real local minimum.

In order to ensure the convergence, a damping factor D is introduced, and the term

$$D\sum_{j=1}^{n}\left(x_j\text{-}x_{j0}\right)^2 \tag{11}$$

is added to the merit function

$$\hat{\phi}=\sum_{i=1}^{m}\bar{f}_i\left(x_j\right)^2+D\sum_{j=1}^{n}\left(x_j\text{-}x_{j0}\right)^2. \tag{12}$$

In minimizing this altered merit function, the design-shift distance $\sqrt{\sum_{j=1}^{n}\left(x_j\text{-}x_{j0}\right)^2}$ remains within a reasonable range in which the approximation error is small. If the value of the damping factor D is small, convergence to a solution is not always guaranteed; if it is too large, the step size becomes too small, and it will take much iteration time to reach a solution. Therefore, the value of D should be chosen to fit the geographical features of the merit function for each iterative calculation. As the automatic control of D had been successfully found, this method called DLS has become the standard optimization method. The optimization path in the parameter space is depicted in Figure 2.

The design follows the path from Start to the Local solution via steps 1, 2, 3,… .

The solution thus obtained is a local minimum of the merit function at a location in parameter space near the starting design. Once the solution reaches the minimum, the control method on the damping factor D forces it to become very large, trapping the solution in that minimum. For that reason, the design cannot jump out of the trap and find a better solution. This is the most serious defect of the DLS method.

2.4 Boundary condition

Generally, each design parameter should have a lower and an upper limit:

$$x_{j\min} < x_j < x_{j\max}.$$

If the value of x_j falls outside this range, the amount of the violation Δx_j should be reduced to as small a value as possible. Δx_j is then regarded as an additional aberration, and its squared value is added to the merit function. The weight of this additional aberration must be larger than those for ordinary error functions.

In addition to design parameters, items such as peripheral thickness of lenses and airspaces should have some lower limit, otherwise the designed lens cannot be realized physically. Such boundary conditions can also be handled within the framework of the DLS method.

3 Global explorer

The feature of the GE that enables it to find multiple designs is the 'escape function' [7–9].

3.1 Escape function

Figure 3 depicts the merit function ϕ as a function of a design parameter x.

When the design falls into a local minimum at x_L, an escape function f_E is set up there as an additional error function defined by

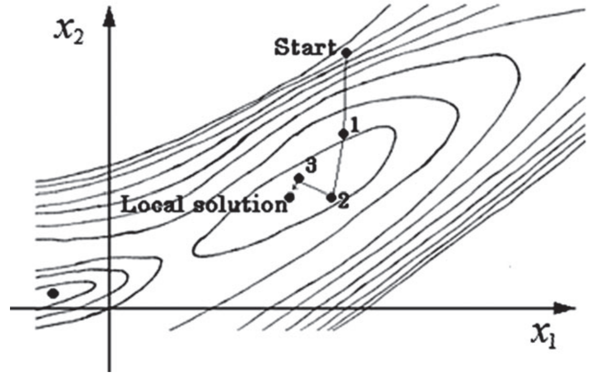

Figure 2 Contour of the merit function in parameter space.

$$f_E = \sqrt{H}\exp\left\{-\frac{1}{2W^2}\sum_{j=1}^{n}\mu_j\left(x_j - x_{jL}\right)^2\right\}, \qquad (14)$$

where x_{jL} is the position of the local minimum from which the design is to escape, and μ_j is the weight for the design parameters, H and W are the height and the width of the escape function, respectively, as illustrated in Figure 4.

The shape of the merit function ϕ around its local minimum changes when the escape function is added. It is raised by an amount of f_E^2, and this enables the design to get out of the local solution and find a new solution if the values of H and W are effectively chosen. Repeating this process, one can automatically find a number of local minima. The contribution of the escape function is depicted in Figure 4 for two design parameter dimensions. (The real design-parameter space is, of course, multidimensional and, hence, impossible to depict graphically.)

In actual cases, where the number of design parameters is large, the problem of choosing appropriate values of H and W is not too delicate. Instead, even crude choices

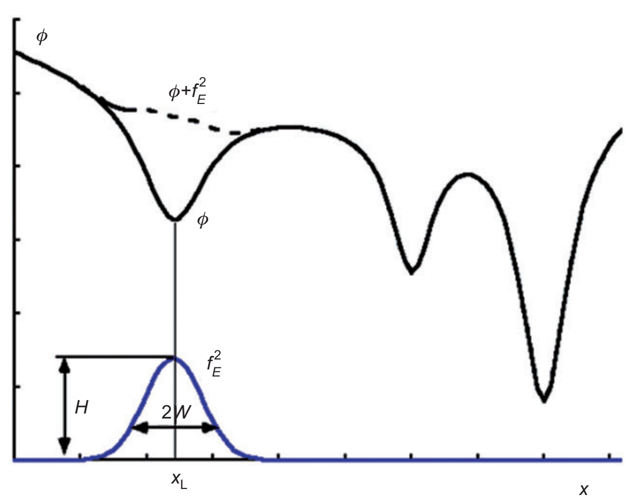

Figure 3 Illustration of local minima and an escape function.

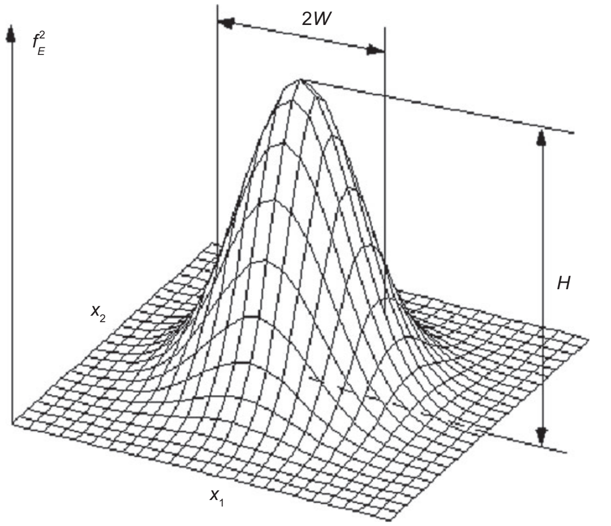

Figure 4 Contribution of the escape function (f_E^2).

generally work well. If escape from the local minimum does not occur, the GE program automatically changes the values of H and W according to a predetermined rule. This process is then repeated until the design escapes from its local minimum.

3.2 Flow chart of GE

Figure 5 depicts a flow chart of the GE program.

The individual steps of the GE program consist of:

1. When the design falls into a local minimum at x_{jL}, the program automatically sets up an escape function with initial values for H and W.
2. DLS optimization is performed for the merit function including the escape function for a few cycles (e.g., 10).
3. DLS optimization is performed again after removing the escape function; the solution thus obtained would hopefully be another local minimum of the merit function.
4. If the newly found solution is not identical with any of those already found, the escape is regarded as a success, and the solution is saved in a lens file.
5. If the escape is not successful, the two parameters H and W are changed according to a predetermined rule (described below), and the processes from (2) to (4) is repeated until a new solution is found.
6. Steps (1)–(5) are repeated until the number of lens solutions filed reaches the requested value.

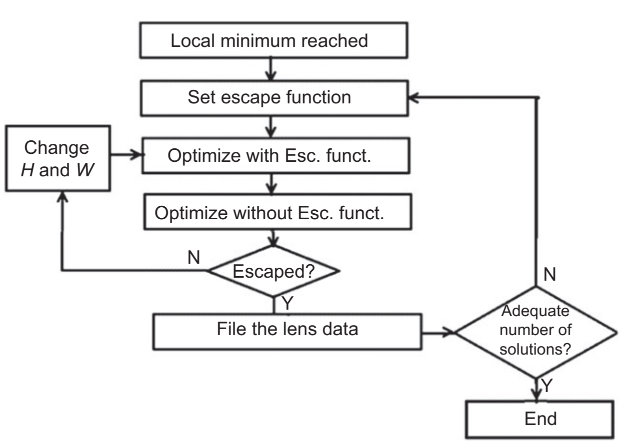

Figure 5 Flow chart of GE program.

In step (4), judgment must be made whether or not the newly found solution is different from the initial design as well as from each of the already filed designs. For this purpose, the distance between two solutions is defined as

$$D_p=\sqrt{\sum_{j=1}^{n}\frac{1}{n}\mu_j\left(x_j\text{-}x_j{'}\right)^2}, \tag{15}$$

where x_j and x_j' are the positions of two solutions in the parameter space. If the newly found solution is separated from any of the previously filed solutions by more than a given threshold distance D_t, i.e.,

$$D_p>D_t, \tag{16}$$

these two solutions are regarded as different, and the escape was done successfully. Otherwise, the escape is judged as unsuccessful.

In applying GE to the lens design, the following parameters should be assigned:

(1) μ_j: the weighting for the j-th design variable. There is no hard and fast rule for determining this value. Some common recommendations are:
   - 1000 for surface curvatures
   - 1 for axial separations between surfaces
   - 10 for refractive indices
   - 1 for v_d value of the material

(2) H and W: As noted previously, these values can be chosen rather arbitrarily. Some recommended initial values are
   - H_0=0.1
   - W_0=0.5.
   - If the escape is unsuccessful, H and W are changed according to the rule
   - $H_k=H_{k\text{-}1}\times H_{\text{mult}}$
   - $W_k=W_{k\text{-}1}\times W_{\text{mult}}$
   - A recommended value for H_{mult} is 2.0 and for W_{mult}, 1.3.

(3) D_t: Threshold distance for identifying two solutions. The recommended value is 5.

4 Global Explorer2

4.1 Winding string

The experience of many designers during the past 50 years [10, 11] has shown that solutions that have nearly the lowest merit function tend to lie within a narrow string winding and stretching through the multidimensional parameter space, as depicted in Figure 6. This has been our experience after making many runs of global optimization with GE. Needless to say, because we live in a 3-D world, it is impossible for us to picture the actual shape of such a string in multidimensional space.

The merit function has a positive value and some lower limit; it can never reach the ideal value of zero. This is because there are always some trade offs among error functions; reduction in one often causes an increase in another. If the lower limit of ϕ is not acceptable, one has to abandon finding a solution with that lens type and change the lens type by increasing the number of lens elements or introducing aspheric surfaces. On the other hand, if the value of ϕ is acceptable, we have freedom to choose one among many solutions in the winding string. In this group, a small change of the merit function can be regarded as meaningless because it is only an average of many aberrations of the optical system. In that case, it is desirable to select the solution that has the lowest sensitivity to manufacturing errors.

4.2 Sensitivity against manufacturing errors

With the 'tolerances' of a system given by $\delta x_1, \delta x_2, ..., \delta x_n$, the sensitivity to manufacturing errors, S, is defined as

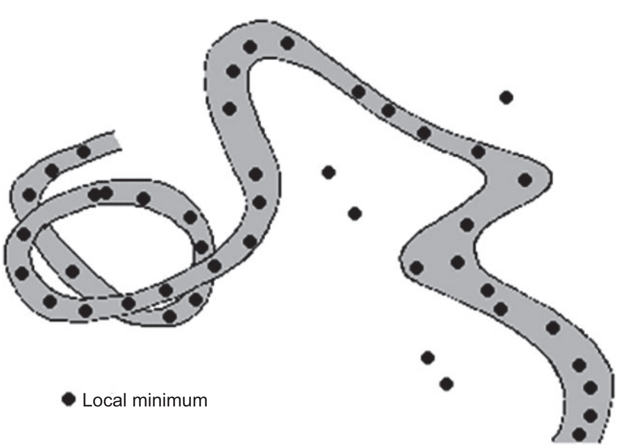

Figure 6 Winding string and local minima.

$$S=\sqrt{\sum_j \left(\frac{\partial \phi}{\partial x_j}\delta x_j\right)^2}, \quad (17)$$

where ϕ is the merit function. In designing lenses, it is important not only to reduce the merit function but also to make S as small as possible. However, an effort to simultaneously reduce ϕ and S using DLS or GE would not be practical due to the extraordinarily long computation time. Instead of using this complicated function S, we employed a root-mean-square value of the incident and refracted angles for some sample rays, denoted as θ

$$\theta=\sqrt{\frac{\sum_{s=1}^{k}\left(i_s^2+r_s^2\right)}{2k}}, \quad (18)$$

where i_s and r_s are the incident and refracted angles of a sample ray at a surface s, respectively. k is the total number of surfaces to be selected for the calculation. These sample rays would typically be the marginal ray to the image center, as well as the upper and lower rim rays to the image corner. Angles at cemented surfaces are excluded. Previous studies have suggested that the value of θ has a good correlation with S, and we can use this value to obtain a robust solution. The advantage of this criterion is that no extra calculation is needed to get it, as those angles have already been calculated in ordinary ray tracing. Moreover, it is not necessary to set up tolerances δx_j for each design parameter. The designer can, in effect, simultaneously improve the performance and the manufacturing robustness of an optical system by controlling the value of θ during the process of optimization. This method is included in GE2.

In developing a high-performance and robust design, the value of θ cannot be too small because the incident and refracted angles have to play a role in image formation. Hence, we apply the condition

$$\theta<\theta_L \ (\theta \text{ segmentation}), \quad (19)$$

under which the merit function ϕ should be minimized [12]. The value of θ_L is set up by the designer. When GE (without θ segmentation) finds a good solution with its sensitivity θ, the designer can set the value of θ_L to be a little smaller. If the solution of GE2 has an acceptable merit function, the designer can further reduce the value of θ_L. This process can be repeated until the merit function increases unacceptably.

4.3 An advantage of GE2

GE2 has the ability to quickly find robust solutions that cannot easily be reached by GE. As an instructive

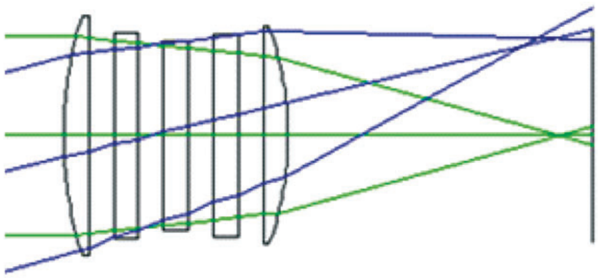

Figure 7 Starting design.

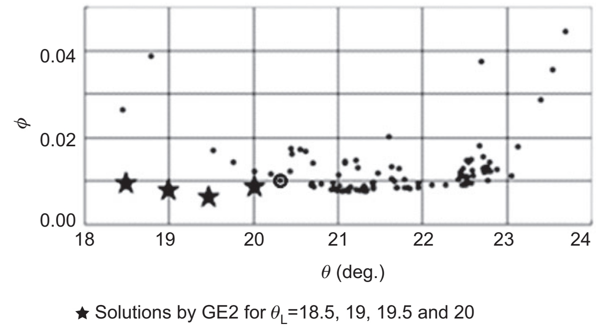

Figure 9 Solutions by GE2.

example, consider the design of a 35-mm camera lens with f=80 mm, F/2, and half angle=15°. The starting design, shown in Figure 7, is composed of five parallel plates with the first and the last surfaces having radii of curvature of 80.000 mm and -65.398 mm respectively, thereby achieving f=80.0 mm. All separations of surfaces are 5 mm, and all glasses are BK7 (Schott).

Naturally, the starting design has very large aberrations. GE is used to find 100 solutions, and these are plotted in Figure 8. Most of the solutions are distributed between θ=20 and θ=23. We can suppose that some forces are acting to prevent the distribution from extending outside this range. We use the colorful language 'Wind 1' and 'Wind 2' to identify these imaginary forces, as depicted in Figure 8. Wind 1 may come from the need of the image-forming power, and wind 2 probably arises to avoid solutions with complicated higher-order aberrations.

If ϕ<0.1 is acceptable, the preferred solution among the 100 would be the one encircled because it has the smallest sensitivity θ.

GE2 can stop the effect of Wind 1 by setting the value of θ_L smaller than that of the encircled solution. More favorable solutions can then be found, as shown in Figure 9.

5 Optimization of aspheric systems

In designing systems containing aspheric lenses, one should limit the number of aspheric polynomial terms. In most cases, an aspheric surface can be expressed by

$$z=\frac{ch^2}{\sqrt{1-(k+1)ch^2}}+\sum_{i=1}^{2m} a_{2i}h^{2i}, \tag{20}$$

where z is the distance from a point on the aspheric surface to a plane that is perpendicular to an axis drawn from the vertex of the asphere. The distance z is thus a function of the height h, as illustrated in Figure 10.

The first term of the above formula describes a conic section, where c is the curvature of the surface at the vertex, and k determines the shape of the conic section. The second term is a polynomial using power coefficients a_{2i}. Any curve can be expressed by this formula with very small error. However, this error generally takes on a ripple-like shape, causing unfavorable effect to the image quality. Thus, one should make the number of polynomial terms as small as possible; we suggest that the number be 3, i.e., the terms a_4, a_6, and a_8. Reducing the number

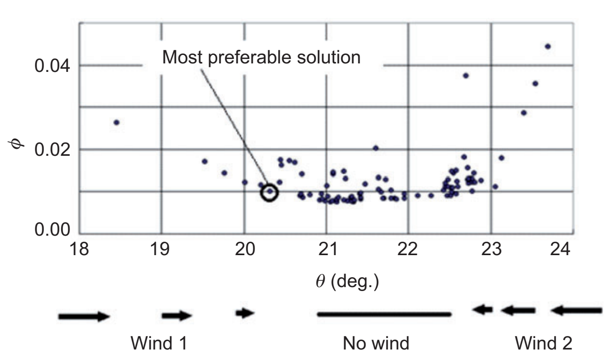

Figure 8 Plot of 100 solutions as a ϕ vs. θ graph.

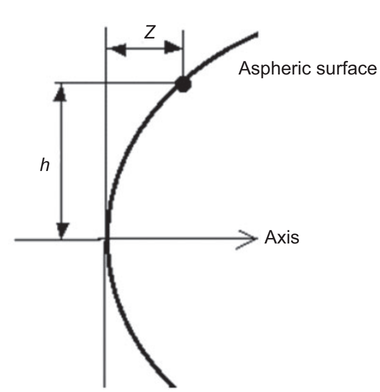

Figure 10 Aspheric surface.

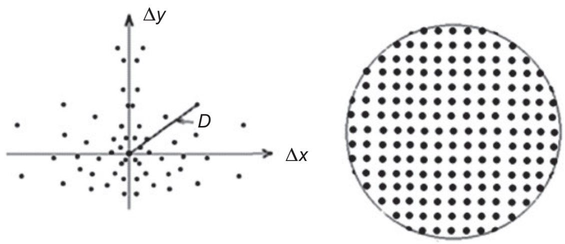

Figure 11 Spot diagram of ray intersections at the image plane (left) and at the entrance pupil (right).

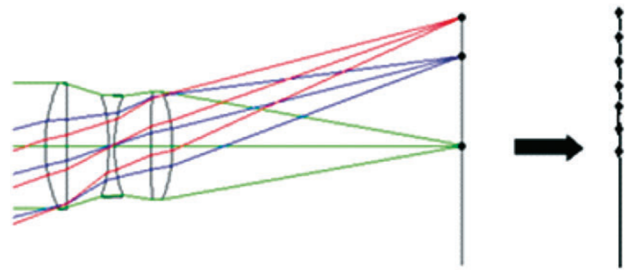

Figure 13 Field points of a lens.

of terms naturally reduces the range of the design parameter space. However, in most cases, good solutions will be found in that subspace because a winding string would most likely extend into that subspace.

The sharpness of an image point can be evaluated by a spot diagram, such as those illustrated in Figure 11, left, which displays the ray intersection points in the image plane of the rays that enter the entrance pupil uniformly, as illustrated in Figure 11, right. The sharpness of the image point is defined as the root-mean-square value of D, the distance between each spot and the center of gravity of the spot distribution.

In designing spherical lenses, it is permissible to trace only a limited number of sample rays, such as those illustrated in Figure 12. By assigning an appropriate weight to each ray, the image point sharpness can be calculated with good accuracy [13]. The reason such limited sampling is permissible for spherical systems is that the aberration theory dictates that the aberration curves

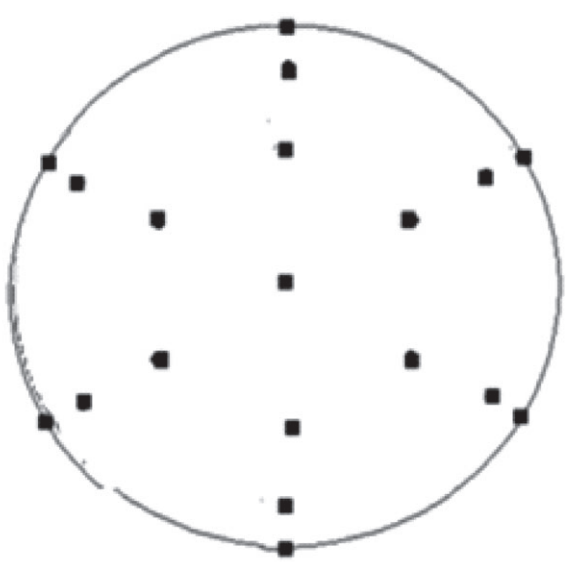

Figure 12 Sampling points in the entrance pupil in the development of a system that uses only spherical lens elements.

are rather smooth. In aspheric systems, there is no such rule, and hence, one is advised to increase the number of rays and to uniformly distribute them over the entrance pupil. The situation is similar for selecting the number of field points. Figure 13 is an example of a triplet lens. For spherical systems, three field points are sufficient, but this number must be increased somewhat in evaluating aspheric systems.

In aspheric systems, aberration curves are not smooth, and the classical aberration theory is not applicable. Some aberrations could be quite harmful even if the merit function is relatively small. Therefore, each lateral aberration should have some upper limit L as shown by

$$|\text{lateral aberration}|<L. \tag{21}$$

The appropriate value of L can be determined by trial and error, in much the same way as described in the θ segmentation method. At the present time, such an interactive design process is the most efficient.

As an illustration of the discussion above, consider the design of a lens for a cellular phone camera with f=10 mm, F/2.8, and a half angle of 31°. All the surfaces are aspheres, each of which has only three polynomial terms. The starting design consists of plane paralell plates of glass BK7 (Schott) as shown in Figure 14.

First, the optimization method with GE2 was applied without L segmentation; this proved to be extremely inefficient. After searching among more than 100 solutions, the best solution found was that shown in Figure 15, Lens A.

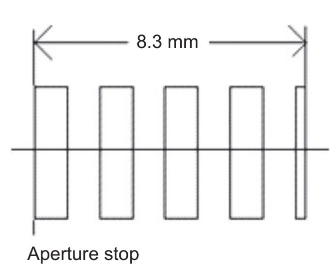

Figure 14 Starting design.

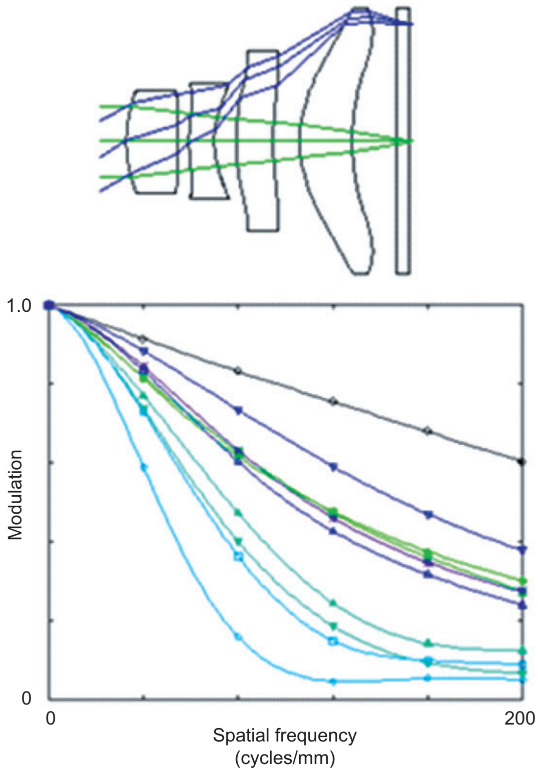

Figure 15 Lens A and its MTF curves for several image heights. The upper curve shows MTF of the aberration free lens.

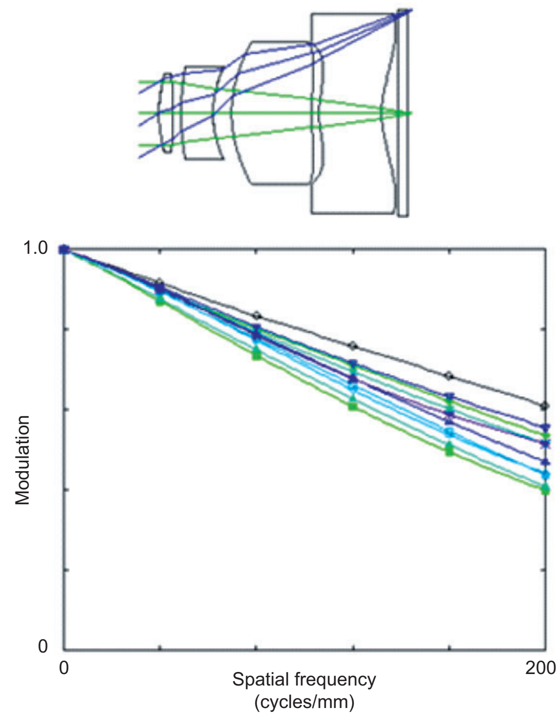

Figure 16 Lens B and its MTF curves for several image heights. The upper curve shows MTF of the aberration-free lens.

When GE2 with *L* segmentation was applied, the algorithm could easily find a good solution (lens B) shown in Figure 16. The reason for this high efficiency comes from *L* segmentation's ability to avoid wasting time in walking around unacceptable solutions.

6 Summary

The optimization program GE employs the well-developed DLS technique. Extensive experience in designing lens systems with GE indicates the existence of 'winding strings' that extend through the design-parameter space. As a result, one can generally select from a group of acceptable designs, and the most robust one among them can be selected using the tactics of θ segmentation. A new version of GE, GE2, was developed to find a good solution that is also robust to fabrication errors. This proved to be highly efficient because one avoids wasting time on solutions that would be unacceptably sensitive to fabrication errors.

The design of aspheric systems is far more complex than the design of spherical systems because the shapes of aberration curves differ markedly. As a result, we suggest the idea of controlling all the lateral aberrations as well as increasing the number of sample rays and field angles. An example is shown for a photographic lens using heavily aspheric surfaces.

The experiments on designing aspherical systems were made by manually adding operands to the ordinarily generated operands' group. For that, we had to do the laborious handiwork. The experiment was conducted only to ascertain the practical advantage of our idea, but we hope that this would be included in the structure of the optimization routine so that the program would be much more friendly to lens designers.

GE is available in the commercial software OSLO [14], and GE2 can be accessed using a macro language of OSLO.

Acknowledgment: The author wishes to express his thanks to Dr. Henry I. Smith (MIT) for his kind suggestions and comments in preparing this manuscript.

Received September 25, 2012; accepted October 23, 2012

References

[1] G. Forbes and A. Jones, Proc. SPIE. 1354, 144–153 (1990).
[2] A. Jones and G. Forbes, in 'Proc. of the IODC', June 6–9, 1994, Rochester, NY, 22, 42–45 (2008).
[3] T. Kuper and T. I. Harris, in 'Proc. SPIE', September 14–18, 1992, Berlin, 1780, 14–28 (1992).
[4] T. Kuper, T. I. Harris and R. Hilbert, 'OSA Proc. of the IODC', June 6–9, 1994, Rochester, NY, 22, 46–51 (1994).
[5] T. Kuper and T. I. Harris, Photonics Spectra, 151–160 (1992).
[6] K. F. Moore, Proc. SPIE 3780, 40–47 (1999).
[7] M. Isshiki, H. Ono and S. Nakadate, "Opt. Rev. 2, 47–51 (1995).
[8] M. Isshiki, H. Ono, K. Hiraga, J. Ishikawa and S. Nakadate, Opt. Rev. 2, 463–470 (1995).
[9] A. Yabe, SPIE 3482, 122–125 (1998).
[10] E. Glatzel and R. Wilson, Appl. Opt. 7, 265–276 (1968).
[11] D. Sturlesi and D. C. O'Shea, Proc. SPIE 1354, 54–58 (1990).
[12] M. Isshiki, D. C. Sinclair and S. Kaneko, Proc. SPIE/OSA 6342, 63420N-2 (2006).
[13] G. W. Forbes, J. Opt. Soc. Am. A5, 1943–1956 (1988).
[14] OSLO: Lens design program available from Lambda Research Corp., 25 Porter Road, Littleton, MA 01460-1434, USA.

Masaki Isshiki is optical consultant belonging to Isshiki Optics. Born in Japan and graduated from the Tokyo University in 1950 he worked for Nikon Corporation in Tokyo (1950–1988), engaged in the field of optical design and optical applications, after that he worked as a professor at Tokyo Polytechnic University (1988–2000) and established Isshiki Optics in 1998. His current research interests are optics design method especially in the development of the optimization.