Reprinted from SPIE Vol. 399—Optical System Design, Analysis, and Production

Improved Convergence with the pseudo-second-derivative (PSD) Optimization Method

Donald C. Dilworth

Optical Systems Design, Inc.
3 Johnson Ave., Medford, Massachusetts 02155

Abstract

The effect of altering the magnitude of a stabilizing factor used in the PSD lens design optimization technique is discussed. Analysis of this effect leads to a method of accounting for the influence of the inhomogeneous second partial derivatives. The improved convergence rate of this method will be demonstrated using examples obtained with the SYNOPSYS lens design program.

Introduction

Most lens design computer programs today utilize the well-known damped-least-squares (DLS) optimization method, or one of numerous variations which have been proposed to yield improved convergence.[1-3] We have previously reported[4] a technique which we call the psuedo-second-derivative (PSD) method, which in some cases accelerates the convergence rate by a factor of ten or more. This method has now been further developed.

The lens design problem requires finding a set of design variables, X, such that a merit function, φ, is minimized. The DLS solution finds a stationary point by iteration, as follows:

Let

$$\varphi = \sum f_i^2 \tag{1}$$

where the f_i are weighted aberrations or other defects to be controlled. To minimize φ we define

$$G_j = 1/2\, \partial\varphi/\partial x_j, \tag{2}$$

and

$$L_{jk} = \partial G_j/\partial x_k, \tag{3}$$

and solve

$$0 = G_j + L_{jk}\Delta x_k \tag{4}$$

for

$$\Delta x_j = -L_{jk}^{-1}\, G_k\ . \tag{5}$$

Δx_j gives the change required in each variable to reduce the value of φ.

Ideally, the elements of the matrix L should be given by

$$L_{jk} = \sum \frac{\partial f_i}{\partial x_j}\frac{\partial f_i}{\partial x_k} + \sum f_i \frac{\partial^2 f_i}{\partial x_j \partial x_k}\ , \tag{6}$$

which involves second partial derivatives as well as first. Owing to the great complexity of computation, no optimization program has fully utilized the mixed partials. Indeed, most techniques in wide use embody one or more assumptions about the latter, and the success or failure of these techniques is due in part to the relative validity of the implied assumptions.

For example, the basic DLS procedure ignores the mixed partials formally, but replaces them with a so-called damping factor which is added to the diagonal of the matrix L.

Although frequently viewed as a device to minimize the variable change while at the same time minimizing the merit function, it is clear that this damping factor would accurately describe the second derivatives if all of the homogeneous second partials were equal, and all mixed partials were zero.

Another common procedure is to assign different but somewhat arbitrary values to the damping that is to be applied to the several types of variables, such as radii and thicknesses. Here we see another assumption, perhaps more valid than the above: that the homogeneous partials of like kind are all equal, and mixed partials are again zero.

Meiron[5] advocates replacing the right-most quantity of equation (6) with the square of the first partials, and scaling the resulting diagonal matrix by an additional damping factor. The assumption in this case is clear, but without obvious theoretical justification.

The PSD method, previously reported, makes the assumption that

$$\frac{\partial^2 f_i}{\partial x_j^2} \approx \frac{\left.\frac{\partial f_i}{\partial x_j}\right|_{\Delta x_j} - \frac{\partial f_i}{\partial x_j}}{\Delta x_j + \epsilon} \quad , \tag{7}$$

which would give accurate homogeneous second partials if higher-order terms as well as the mixed terms were zero.

Progress in optimization so far has been obtained by increasing the accuracy of the implied assumptions rather than by any increase in actual computation effort or by obtaining additional information about the lens. We now continue this trend. A recent improvement to the PSD method involves a more realistic assumption about the mixed terms, that they are not necessarily all zero.

Improved PSD Method

It was quickly realized that the PSD method, as defined by equation (7), required an extra term in the denominator, ϵ , originally a stabilizing factor to prevent instability in the approximate second derivatives: if the value of ΔX were to be extremely small at a particular iteration, as sometimes happens, the approximation would give a very large value of the derivative causing the next ΔX to be even smaller, and so on. An arbitrary value of .001 for ϵ seemed to give satisfactory convergence without this instability, in systems on which it was tested.

We noticed that a larger value, say .1, often gave far greater improvement at the start of a run, but then quickly stagnated at a point short of the true optimum -- while a very small value often worked poorly at first, but then gave outstanding convergence at the end.

Clearly, the stabilizing factor ϵ is required because the mixed partial second derivatives, which are ignored in equation (7), have a measurable influence on the approximation. This influence may be accounted for in the following way.

Instead of the approximation in equation (7), let us write

$$\left(\left.\frac{\partial f_i}{\partial x_j}\right|_{\Delta x_j} - \frac{\partial f_i}{\partial x_j} \right) \approx \frac{\partial^2 f_i}{\partial x_j^2} \Delta x_j + \underbrace{\underbrace{\sum_k \underbrace{\frac{\partial^2 f_i}{\partial x_j \partial x_k}}_{S} \Delta x_k}_{B}}_{Q} \Bigg|_{k \neq j} \tag{8}$$

where the quantity S is the (unknown) true mixed partial derivative, whose effect we wish to determine.

Since the purpose of an acceleration technique is to reduce the amount of computer time required to optimize a lens, we prefer to make some kind of assumption rather than actually compute the value of S, which would be very time consuming. This assumption need not be exact; the previous methods assumed that the mixed terms were zero, which they probably are not, and yet these methods work fairly well -- all we need is a better assumption.

Mixed Partial Approximation

Without direct knowledge, it makes no sense to discuss individual terms. But perhaps we can deal with the mixed partials in a statistical sense. It seems plausable to describe the mixed second derivatives, S, in the following way:

1. Their values are randomly distributed.

2. The mean value is approximately zero.

3. The standard deviation, σ, is approximately equal to the value of the homogeneous second partial.

These assumptions merely state that we would expect the mixed partials to be more or less of the same magnitude as the homogeneous partials, but with random variations in sign and magnitude. In such a situation, they may be accounted for.

We have

$$B_{ijk} = S_{ijk}\Delta x_k \tag{9}$$

and

$$\sigma_B = \sigma_S \Delta x_k \tag{10}$$

which gives the standard deviation in the quantity B.

Since

$$Q_{ij} = \sum_k B_{ijk} \quad , \tag{11}$$

we can write the variance of Q as

$$\sigma_Q^2 = \sum_k \sigma_B^2 \tag{12}$$

and the standard deviation of Q is therefore

$$\sigma_Q = \sqrt{\sum_k \sigma_S^2 \Delta x_k^2} \tag{13}$$

$$= \sqrt{\sum_k \left(\frac{\partial^2 f_i}{\partial x_j^2}\Delta x_k\right)^2} \tag{14}$$

therefore,

$$\sigma_Q = \frac{\partial^2 f_i}{\partial x_j^2}\sqrt{\sum_k \Delta x_k^2 \Big|_{k\neq j}} \quad . \tag{15}$$

If we replace the unknown value of Q_{ij} in equation (8) with its expected value,

$$\left(\frac{\partial f_i}{\partial x_j}\bigg|_{\Delta x_j} - \frac{\partial f_i}{\partial x_j}\right) \approx \frac{\partial^2 f_i}{\partial x_j^2}\,\Delta x_j + \frac{\partial^2 f_i}{\partial x_j^2}\sqrt{\sum_k \Delta x_k^2}\;\bigg|_{k\neq j} \qquad (16)$$

we can solve for the approximate homogeneous second derivative:

$$\frac{\partial^2 f_i}{\partial x_j^{\,2}} \approx \frac{\frac{\partial f_i}{\partial x_j}\Big|_{\Delta x_j} - \frac{\partial f_i}{\partial x_j}}{|\Delta x_j| + \sqrt{\sum_k \Delta x_k^2}\;\Big|_{k\neq j}} \; . \qquad (17)$$

Here we see that the stabilizing factor, ϵ has been replaced with a term in which the effects of the mixed partials appear, on a statistical basis. It is immediately clear that at the early stages of a design, when the variables are undergoing large changes, the quantity under the radical in equation (17) is large -- while at the end of the design, when the variables are changing very little, that quantity is very small. When compared with the most effective values of ϵ that we found earlier, it seems that this relationship is exactly what is required to give the fastest convergence at all stages of a run.

Equation (17) includes the absolute value of Δx rather than Δx itself, for the following reason: there is no sign information in the radical, and we want the approximate second derivative to be reduced in magnitude when the radical is large (because then more of the observed change in first derivatives is due to the mixed partials). The absolute value sign assures that this will be the case. Since there is no sign information on the second derivative after this step, it also makes sense to take the absolute value of the quantity on the right side of equation (6); since we are looking for minima of the merit function, a large second derivative should reduce the solution vector rather than lengthen it, and this occurs with a positive contribution at that point.

Effectiveness of the Method

Having established a plausible, if not rigorous justification for the new PSD approximation, the crucial test is to determine what happens when the method is used on a variety of lens design problems.

In the previous description of the PSD method, we showed a highly contrived problem in which various degrees of nonlinearity could be assigned to two aberrations which were to be corrected by two variables. The system, which was labeled "case D" in the earlier paper is shown in Fig. 1.

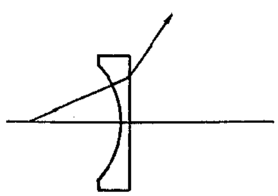

Figure 1. Test case "D"

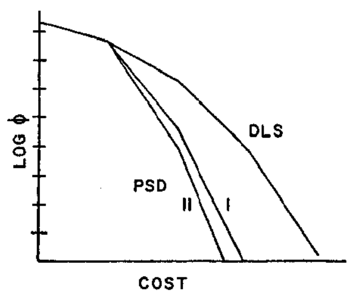

Figure 2. Convergence of case "D" at 20 degrees

The two variables were the lens thickness and the entrance pupil diameter, and the two aberrations were the output marginal ray angle and the sum of the lens thickness and surface sag at the margin. The first aberration may be highly nonlinear with respect to the aperture, while the second is only slightly nonlinear. The thickness is exactly linear in all cases.

We tested the new PSD method for this case and compared the rate of convergence with that from the old PSD method and the basic DLS technique. We would expect that if the degree of nonlinearity were similar for both variables, as would be the case when the output ray angle is small, the DLS method should be as good as any. On the other hand, if a large angle were requested, the DLS method should fare poorly while the PSD methods should succeed. Our results confirm this expectation.

Fig. 2 shows the rates of convergence of the three methods for the case when the output ray angle was 20 degrees, and Fig. 3 the case of 89.9 degrees. In the latter case DLS was unable to change the system while the PSD method experienced no difficulty. The new approximation converged somewhat faster than the old.

A second example involved optimizing a very fast mirror with a conic constant and three aspheric terms. The DLS method frequently has trouble with power-series aspherics, since the nonlinearities involved vary widely. The convergence rates for this example are shown in Fig. 4. Here the DLS method performed rather well, about the same as the old PSD method, but in this case the new method is clearly superior.

A third example is shown in Fig. 5. It is a three-element germanium beam expander lens which would not appear to be a difficult design since the power and angles are all modest. However, the DLS method was nearly powerless on this design, as shown in Fig. 6. Both PSD methods worked well, with the new one arriving at the minimum first. Examination of the printed diagonal terms from the PSD runs shows that surface 5 in this lens is more linear by a factor of a thousand or so than the other variables, and that reversing the bending of the third element was essential to the success of the design. Apparently, the DLS method could not change surface 5 rapidly enough to arrive at the better lens shape since the damping that was employed was far higher than the second derivative term for that variable.

For a fourth example we ran the triplet shown in Fig. 7. There are two solution regions for this lens, and the starting configuration resembled the poorer minimum more than the better. The convergence curves for this lens are shown in Fig. 8. The DLS method slowly crept toward the poorer local minimum, as did the PSD methods for about half of the run. But then they turned and abruptly sped toward the better solution. This case deserves more study, since it suggests that what would be a local minimum to a DLS program might not appear so to a program with better second-derivative information.

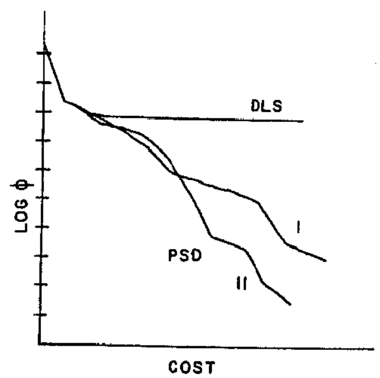

Figure 3. Case "D" at 89.9 degrees

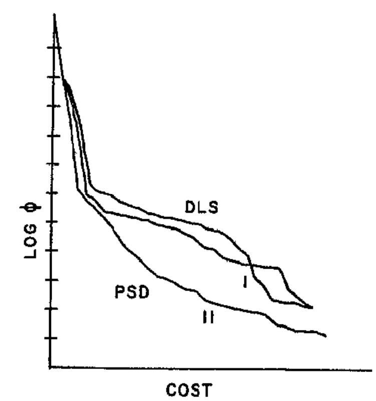

Figure 4. Convergence of aspheric mirror

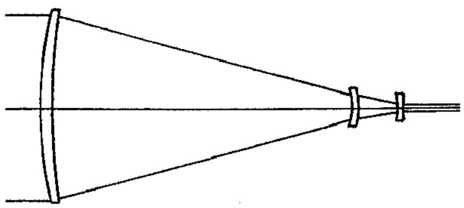

Figure 5. Afocal example

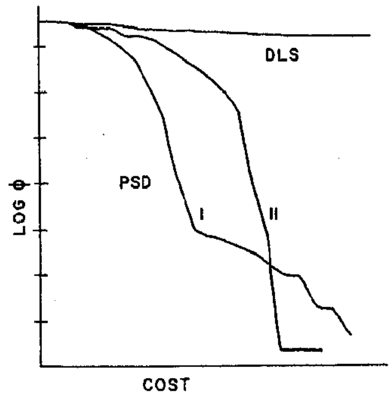

Figure 6. Convergence rate of afocal example

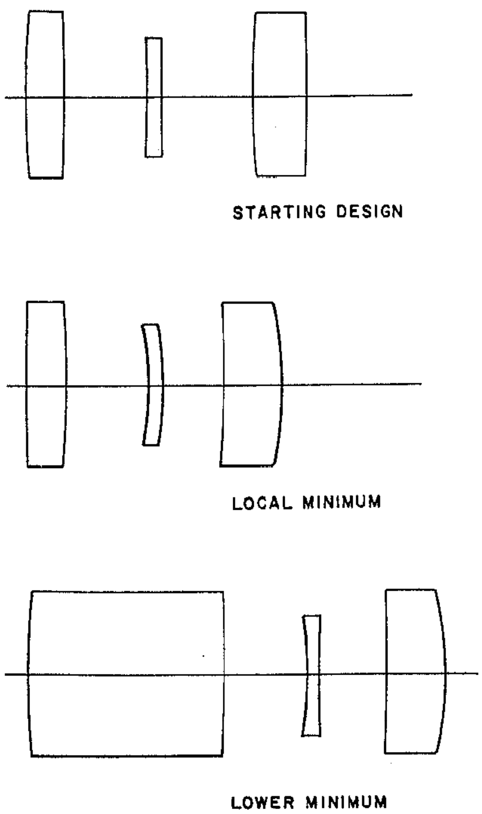

Figure 7. Triplet example

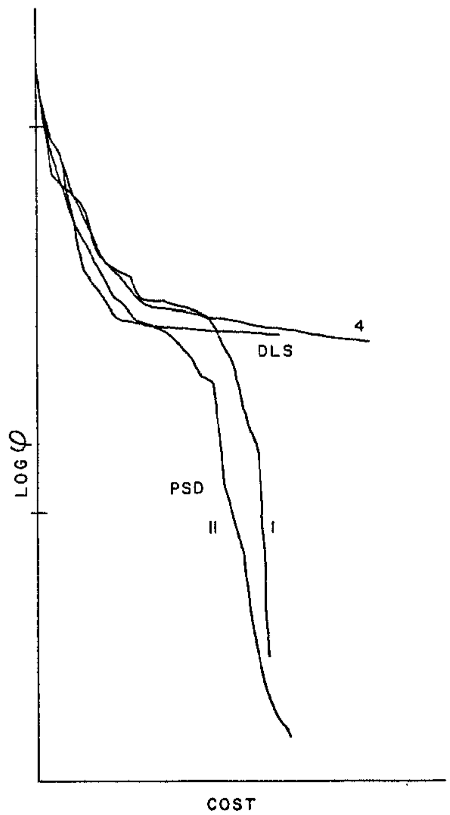

Figure 8. Convergence rate of triplet

Other Variations of the PSD Method

There are further modifications to the PSD method which may be of interest. An idea suggested by Huber[6] is to utilize the approximate second derivative information to predict the values of the first derivatives and even the merit function after altering the design each iteration. If accurate, the prediction would enable one to bypass the time-consuming calculation of these quantities after each pass. Although this method is reported to be of benefit in a DLS optimization, we have found in every case with the PSD method that the reduction in computer time is about proportional to the reduction in improvement due to the reduced accuracy of the extrapolated quantities. So there is no net gain from this approach.

Another idea, described by Faggiano[7] is to scale the additions to the diagonal matrix by an additional damping factor, which starts with a very low number and is increased only if necessary.

All versions of the optimizing program which we have used have employed a scaling factor, as previously noted[4]. A further refinement is to explore the effect of this scaling factor at each pass so as to find the optimum value before accepting the solution and recomputing the derivatives. Here the results are somewhat unexpected. The improvement per iteration is usually greater if this search is performed, but the improvement per unit of computer time is usually slightly poorer than if we merely accept the first value that gives an improved lens and do not fit a parabola to the data and find the best.

An even more surprising result is shown in Fig. 8. This triplet was run with the new PSD method with a full scaling factor search and optimization, and the convergence curve is labeled "4". Even though the PSD method is ordinarily capable of finding the lower minimum for this lens, the scaling search has somehow prevented it in this case, and the solution is no better than that from the ordinary DLS method. It appears that finding the maximum improvement at each pass is of no consequence in the long run, and in some cases may actually impede the optimization.

Yet another idea, described by Robb[8], is to scale the solution vector ΔX so as to maximize the improvement each pass. We have found that the best scaling factor for a PSD solution is nearly unity in every case, and there is little advantage to exploring in this dimension. These conclusions apply, of course, only to the the PSD algorithm, and any of these schemes may be of value in an ordinary DLS program.

Conclusions

We have developed an improved lens optimization algorithm which in many cases converges faster than the previous PSD method and nearly always faster than the ordinary DLS method. The statistical assumptions that are inherent in this technique regarding the inhomogeneous second partial derivatives therefore appear for the most part to be justified.

Because of its superior convergence rate, this technique has been incorporated into the SYNOPSYS optical design program as a default mode of operation. Users of SYNOPSYS now have a choice of three optimization algorithms, including ordinary DLS and the old and new PSD methods; it now appears that the new PSD method is the superior method for most lenses, with the option to optimize the scaling factor turned off.

References

1. D. R. Buchele, Appl. Opt. 7, 2433 (1968).
2. D. P. Feder, Appl. Opt. 2, 1209 (1963).
3. C. G. Wynne and P. M. J. H. Wormell, Appl. Opt. 2, 1233 (1963).
4. D. C. Dilworth, Appl. Opt. 17, 3372, (1978).
5. J. Meiron, J. Opt. Soc. Am. 55, 1105 (1965).
6. E. D. Huber, Appl. Opt. 21, 1705 (1982).
7. A. Faggiano, Appl. Opt. 19, 4226 (1980).
8. P. N. Robb, Appl. Opt. 18, 4191 (1979).