Performance Comparison between ASA and DIRECT

Hong Zhu and David B. Bogy

Computer Mechanics Laboratory

Department of Mechanical Engineering

University of California at Berkeley

Berkeley, CA 94720

ABSTRACT

This report compares two fundamentally different global optimization techniques: the stochastic ASA (Adaptive Simulated Annealing) algorithm and the deterministic DIRECT (DIviding RECTtangles) algorithm. Both algorithms are powerful tools for dealing with strongly non-linear problems and they are quite robust. However, through numerical simulations as well as slider ABS optimization case studies, DIRECT is found to be more convenient for practical usage than ASA. Besides, due to its much faster convergence rate, DIRECT can find the global minimum more quickly than ASA. Also it is quite easy for DIRECT to handle the manufacturing tolerance issue. In summary, we found the DIRECT algorithm clearly outperforms the ASA algorithm in our test cases, and we consider DIRECT more suitable for slider ABS optimization than is ASA.

1. INTRODUCTION

We introduced the Adaptive Simulated Annealing (ASA) algorithm and the DIRECT algorithm in the CML technical reports 00-010 and 01-003, respectively. As discussed previously, the ASA algorithm and the DIRECT algorithm are both global optimization techniques. However, these two algorithms belong to two fundamentally different categories. ASA is a stochastic algorithm, which means a random process is introduced within the algorithm. DIRECT is a deterministic algorithm, and thus no random process is involved.

The ASA algorithm, developed by Ingber in 1989, is one of the latest developments of the Simulated Annealing (SA) family, and it has been applied to many areas. Research on the Simulated Annealing technique has also become intensified in recent years. On the other hand, the DIRECT algorithm, developed by Jones in 1993, is a relatively new algorithm and has not been widely used.

Both algorithms are powerful tools for dealing with strongly non-linear problems. Neither requires a definite form of the objective function, but instead they both treat the solver and the objective function as "black boxes". After generating the samples (input) and sending them to the solver, only the values of the objective function given by the solver (output) are needed. For both algorithms, only continuity is required for the optimization problems, i.e., the objective function must be continuous within the search space.

Both algorithms are quite robust. Although ASA has the highest convergence rate within the SA family, it is still relatively slow. It generally needs to be "tuned up" (Ingber, 1996) before it can be applied successfully to a specific problem. Also, since ASA is a stochastic algorithm, it's reasonable to expect different optimization results for a limited number of samples generated if there are any changes in the initial conditions or algorithm parameters. However, statistically, ASA will give the same results if a large number of samples are generated. DIRECT is a deterministic algorithm. Therefore, there are no random factors involved in its optimization process. Interestingly, DIRECT does not need to be "tuned up" for a specific problem, because there are almost no parameters that need to be changed. Jones (1993) and Gablonsky et al. (1998, 2000) showed that DIRECT has a very fast convergence rate. Thus, it can find the global minimum very quickly as compared with other algorithms.

Another distinction between these two algorithms is their relative difficulty of implementation. We found that ASA is much easier to implement than DIRECT. However, once the more difficult implementation of DIRECT is finished, the effort will be rewarded by its amazingly high efficiency.

Next we will compare the relative performance of ASA and DIRECT using numerical simulation and the slider ABS optimization case study.

2. PERFORMANCE COMPARISON BY NUMERICAL EXPERIMENTS

2.1 Easy test function cases

The test functions used here include 2-D, 10-D and 20-D functions. These functions have only one global and local minimum point, and the minimum values of these functions are zero. They are defined as follows:

2-D: $F(x_1, x_2) = (x_1 - 0.4)^2 + (x_2 - 0.2)^2.$

10-D: $F(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}) = (x_1 - 0.1)^2 + (x_2 - 0.2)^2 + (x_3 - 0.3)^2 + (x_4 - 0.4)^2 + (x_5 - 0.5)^2 + (x_6 - 0.6)^2 + (x_7 - 0.7)^2 + (x_8 - 0.8)^2 + (x_9 - 0.9)^2 + (x_{10} - 1.0)^2.$

20-D: $F(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}) = (x_1 - 0.05)^2 + (x_2 - 0.1)^2 + (x_3 - 0.15)^2 + (x_4 - 0.2)^2 + (x_5 - 0.25)^2 + (x_6 - 0.3)^2 + (x_7 - 0.35)^2 + (x_8 - 0.4)^2 + (x_9 - 0.45)^2 + (x_{10} - 0.5)^2 + (x_{11} - 0.55)^2 + (x_{12} - 0.6)^2 + (x_{13} - 0.65)^2 + (x_{14} - 0.7)^2 + (x_{15} - 0.75)^2 + (x_{16} - 0.8)^2 + (x_{17} - 0.85)^2 + (x_{18} - 0.9)^2 + (x_{19} - 0.95)^2 + (x_{20} - 1.0)^2.$

For all these cases, $x_i \in [0,1], \; i = 1, \ldots 20.$

As the dimensions of a problem increase, the search space becomes much larger, and the difference between the convergence rates of ASA and DIRECT become more obvious. Figures 1, 2 and 3 show the comparison between convergence rate for ASA and DIRECT for the 2-D, 10-D and 20-D cases, respectively.

Figure 1 shows that, for the 2-D case, ASA and DIRECT demonstrate similar convergence curves. For the 10-D case shown in Fig. 2, we see that ASA and DIRECT reveal different convergence rates. ASA shows a fast convergence rate at the initial stage, but slows thereafter. However, DIRECT shows a steady and very fast convergence rate throughout the whole optimization process. This phenomenon can be observed more clearly in Fig. 3 for the 20-D case. Figure 3 shows that ASA converges quite quickly when the number of function evaluations is less than 10000, but after that it converges very slowly. ASA required more than 80000 function evaluations to find the best value of 10^{-3}, whereas DIRECT took fewer than 15000 function evaluations to accomplish the same task. Also, the convergence rate of DIRECT remains steady throughout the entire process. The best value found by DIRECT is lower than 10^{-4} when the number of function evaluations reaches 20000. In Fig. 3, we also plot the convergence curve for DIRECT-III, a strongly locally biased variation of the standard DIRECT algorithm introduced in the CML technical report 01-007. It is clear that DIRECT-III converges even faster than DIRECT in this case.

Consider another 2-D test function that is defined as:

$F(x_1, x_2) = x_1^{\,2} + x_2^{\,2} + 0.3\, z \sin(13\pi z)^2,$

where $z = x_1^2 + x_2^2 + 0.0001$ and $x_1, x_2 \in [-5,4]$. If we normalize the range of variables x_1 and x_2 into $[0,1]$, then its global minimum is 5.0039E-10 at (5/9, 5/9). The 2-D contour lines and 3-D surface shape of this function are shown in Figs. 4 and 5, respectively. In Fig. 4 the round dot denotes the location of the global minimum point. From Figs. 4 and 5 it is observed that this function has many local minima due to the perturbation term $0.3\ z\ sin(13\pi z)^2$. However, since the perturbation term contributes very little to the function value, it is still considered here to be an "easy" test function.

Figure 6 shows the convergence comparison between ASA and DIRECT for this 2-D test function. ASA and DIRECT show similar convergence rates in this case.

Figure 7 shows the convergence comparison for ASA for five different initial conditions: (0.5, 0.5), (0, 0), (0, 1), (1, 1) and (1, 0). Although the convergence curves for these five different initial conditions are different, statistically they are similar to each other. In Fig. 7 we see that different initial conditions result in different optimization processes for ASA. However, in this case all of them exhibit similar convergence rates and all of the curves converge to the global minimum.

2.2 Tough testing function cases

The so-called "tough" functions are the ones whose global minima are difficult for the optimization technique to find. Mostly this is caused by either multiple local minima or a wide "flat" area around the global minimum point. These features make the optimization difficult since it's easy for the process to get trapped at a local minimum.

We investigated two tough test functions with one global minimum and multiple local minima. The first function is the 4-D Colville function. It is defined as:

$$F(x_1, x_2, x_3, x_4) = 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1\,((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1),$$

where $x_1, x_2, x_3, x_4 \in [-10, 10]$. If we normalize the range of variables x_1, x_2, x_3 and x_4 into $[0,1]$, then its global minimum point is (0.55, 0.55, 0.55, 0.55) and the global minimum is 0.

Figure 8 shows the convergence comparison between ASA and DIRECT for the 4-D Colville function. ASA shows a very slow convergence rate after 1000 function evaluations, whereas DIRECT shows a very steady and fast convergence rate.

The second function is the local Shubert function, which is defined as follows:

$$F(x_1, x_2) = -(\sum_{i=1}^{5} i \sin((i+1)x_1 + i) + \sum_{j=1}^{5} j \sin((j+1)x_2 + j)) ,$$

where $x_1, x_2 \in [-5, 5]$. It has one global minimum point and 100 local minimum points. If we normalize the range of variables x_1 and x_2 into $[0,1]$, then its global minimum point is (0.4508609, 0.4508609) and the global minimum is -24.062499. The 2-D contour and 3-D surface of the local Shubert function are shown in Figs. 9 and 10, respectively. The round dot in Fig. 9 denotes the global minimum point.

Figure 11 shows the convergence comparison between ASA and DIRECT when ASA takes the midpoint (0.5, 0.5) as the initial point. We see that, for this case, although both algorithms did find the global minimum, ASA found it faster than DIRECT.

However, if we change the initial point for ASA, the results are quite different. To do that, we divided the square search space into an 8 x 8 grid, and used each grid point as the initial point for the optimization using ASA. Consequently, we tested 81 different initial points for ASA, and we show the results in Fig. 12. We combined the contour lines of the local Shubert function with the 8 x 8 grid. The triangular symbols represent the initial points for which ASA failed to find the global minimum. For 18 out of the total 81 initial points ASA failed to find the global minimum. This verifies that the optimization results of ASA are dependent on the initial conditions, which in this case were the initial points. This dependence is caused by the combined effects of the ASA algorithm itself and the embedded random process. The DIRECT algorithm always chooses the midpoint of the search space as its initial point and there is no embedded random process. Therefore, there is no such dependence for DIRECT.

Even thought ASA fails to find the global minimum in Fig. 12, this does not mean that ASA is not a global optimization technique. Remember that ASA has many parameters and generally needs to be "tuned up" for each specific problem. In this case, we used the default settings for ASA and did not tune it up. For all those initial points for which ASA failed to find the global minimum, if we adjusted some of the parameters (e.g., setting a higher initial temperature, or simply changing the random number generator), ASA was able to find the global minimum point.

In fact, leaving many parameters to be "tuned up" by users for a specific problem is one of the disadvantages of ASA, especially for problems with expensive function evaluations, because users usually do not have the luxury of "tuning up" the ASA until it performs well. In contrast, the DIRECT algorithm, being deterministic and not needing parameters to be changed by the user, is much more convenient for practical usage.

Note that all of the test functions we chose only have one global minimum. In practical terms it is very difficult for ASA to handle problems with multiple global minima, especially those problems with multiple global minima and many local minima (for example, the Shubert function). We have shown in the CML technical report 01-003 that DIRECT can handle the multiple global minima very well, even for some very "nasty" test functions.

3. PERFORMANCE COMPARISON USING SLIDER ABS OPTIMIZATION

3.1 Air bearing design optimization problems

Here we investigate a 2-D and a 3-D ABS optimization problem. For both cases, we choose the same NSIC 7nm FH slider as the prototype slider. Its rail shape and the 3-dimensional rail geometry are shown in Figs. 13 and 14, respectively.

The slider is a Pico slider (1.25×1.0mm), which flies over a disk rotating at 7200 RPM. Its flying heights are all around 7nm from OD to ID. The optimization goals are also the same in both cases: to lower the flying heights to the target flying height (i.e., 5nm), and at the same time maintain a flat roll profile across the three different radial positions OD, MD and ID. We also define the same objective function for both cases as:

1× *(FH Max Difference term)* + ***9*** × *(FH term)* + ***1*** × *(Roll term)* +

1 × *(Roll Cutoff term)* + ***1*** × *(Pitch Cutoff term)* + ***1*** × *(Vertical Sensitivity term)* +

1 × *(Pitch Sensitivity term)* + ***1*** × *(Roll Sensitivity term)* + ***1*** × *(Negative Force term)*.

Note that since we are primarily concerned with the flying heights, we put a heavier weight (9) on that term. All the objective terms are normalized and their definitions can be found in the CML technical report 01-016.

For the 2-D ABS optimization problem, only two original constraint points are defined, as shown in Fig. 15. For the 3-D ABS optimization problem, three original constraint points are defined, as shown in Fig. 16.

3.2 Optimization results

Using the same initial design, constraints, and objective function, we carried out the optimization for the 2-D and 3-D ABS optimization problems using both ASA and DIRECT. Notice that for the purpose of comparison we only use the standard DIRECT algorithm without manufacturing tolerance or hidden constraints. We use 200 function evaluations for DIRECT in both the 2-D and 3-D cases.

3.2.1 2-D ABS optimization results

Figure 17 shows the convergence comparison between ASA and DIRECT for the 2-D ABS optimization case. Here, the convergence curves of ASA and DIRECT are quite similar. The optimized ABS design found by ASA has the objective function value of 5.561, whereas the optimized ABS design found by DIRECT has the objective function value of 5.571. ASA found a slightly better ABS design in this case. Figures 18 and 19 show the optimized designs found by ASA and DIRECT, respectively. The gray lines in those two figures represent the initial ABS design, and the dark lines represent the optimized ABS design.

Figures 20 and 21 show the optimization results for the 2-D case using DIRECT and ASA, respectively. Here the small dots represent the sample ABS designs generated by the algorithms, and the circular dots represent the best-so-far ABS designs found by the algorithms during the optimization process.

It can be observed from Figs. 20 and 21 that while the optimization results obtained by using the deterministic DIRECT algorithm show a very regular pattern, the optimization results obtained using the stochastic ASA algorithm also show an interesting "band" pattern. That is, most of the sample points generated by ASA for this 2-D case were located in two "bands", which are shown in Fig. 22 by the two rectangles ($a \times b1$ and $b \times a1$) surrounded by the dashed lines.

This "band" pattern can be explained by the following:

Let $a2 = a - a1$ and $b2 = b - b1$, and suppose that for the whole optimization process there is a probability of 70% for the new sample points to be generated within the *a1* and *b1* intervals for the first and second original constraint points, as shown in Fig. 23. Then, for the whole search space, the probability distribution for all the sample points is:

$$((70\%)^{a1} + (30\%)^{a2}) \times ((70\%)^{b1} + (30\%)^{b2})$$

$$= (49\%)^{a1+b1} + (21\%)^{a1+b2} + (21\%)^{a2+b1} + (9\%)^{a2+b2}.$$

We can also verify this by looking at the intersection of the two "bands", which is $(a1+b1)$ and is shown in the shaded area of Fig. 22, the section with the greatest density of points.

3.2.2 3-D ABS optimization results

Figure 24 shows the convergence comparison between ASA and DIRECT for the 3-D ABS optimization case. Here we see that the DIRECT algorithm clearly has a much faster convergence rate than the ASA algorithm. In this case, it takes only about 100 sampling designs for DIRECT to converge to the optimal design, whereas it takes more than 600 sampling designs for ASA.

Also note that the objective function value of the final optimized design obtained by using DIRECT is 4.46. For ASA, the final optimized design's objective function value is 4.74. Since a smaller objective function value means a better design, the DIRECT algorithm obtained a better-optimized design than ASA algorithm.

Figures 25 and 26 show the optimized designs found by ASA and DIRECT, respectively. The gray lines in those two figures represent the initial ABS designs, and the dark lines represent the optimized ABS designs. Figures 27 and 28 show the comparison between the FH and the Roll of the optimized ABS designs found by ASA and DIRECT, respectively. Figures 27 and 28 show that both algorithms found greatly optimized ABS designs, but the optimized ABS design found by DIRECT has even more uniform FHs around the 5nm target FH and flatter rolls. Moreover, DIRECT found the optimized design much faster than did ASA, due to its much faster convergence rate.

4. CONCLUSION

The ASA algorithm is one of the latest developments of the Simulated Annealing (SA) family, and it has been applied to many areas. The DIRECT algorithm is a relatively new algorithm and has not yet been widely used.

The ASA algorithm and the DIRECT algorithm are both global optimization techniques. Both algorithms are powerful tools for dealing with strongly non-linear problems, and neither require a definite form of the objective function. The only requirement for both algorithms is that the objective function must be continuous within the search space.

However, these two algorithms belong to two fundamentally different categories. ASA is a stochastic algorithm, which means it introduces a random process within it. DIRECT is a deterministic algorithm and no random process is involved.

Both ASA and DIRECT are quite robust. ASA generally needs to be “tuned up” in order to be successfully applied to a specific problem. Also, since ASA is a stochastic algorithm, it is reasonable to expect different optimization results for a limited number of samples generated for any changes in initial conditions or algorithm parameters However, statistically, ASA gives the same results if a large number of samples are generated. DIRECT is a deterministic algorithm, and therefore there are no random factors involved in the optimization process. DIRECT does not need to be “tuned up”. Thus, DIRECT is more convenient for practical usage.

Although DIRECT is initially more difficult to implement than is ASA, it also has a much higher convergence rate than dose ASA. Thus, DIRECT can find the global minimum more quickly than ASA. This property is observed more clearly for higher dimensional problems. We have verified this by numerical simulations as well as in slider ABS optimization case studies.

Also notice that it is quite easy for DIRECT to handle the manufacturing tolerance issue. This, together with its high convergence rate, makes DIRECT more suitable for the slider ABS optimization.

In summary, the DIRECT algorithm clearly outperforms the ASA algorithm in our test cases, and we consider it more suitable for the slider ABS optimization than is ASA.

ACKNOWLEDGEMENT

This study is supported by the Computer Mechanics Laboratory (CML) at the University of California at Berkeley and partially supported by the Extremely High Density Recordings (EHDR) project of the National Storage Industry Consortium (NSIC).

REFERENCES

1. Zhu, H. and Bogy, D., 2000, “Optimization of Slider Air Bearing Shapes using Variations of Simulated Annealing”, Technical Report 00-010, Computer Mechanics Laboratory, University of California at Berkeley.

2. Zhu, H. and Bogy, D., 2001, “DIRECT Algorithm and its Application to Slider Air Bearing Surfaces Optimization”, Technical Report 01-003, Computer Mechanics Laboratory, University of California at Berkeley.

3. Zhu, H. and Bogy, D., 2001, “Locally Biased Variations of the DIRECT Algorithm and their application to the Slider Air Bearing Surfaces Optimization”, Technical Report 01-007, Computer Mechanics Laboratory, University of California at Berkeley.

4. Zhu, H. and Bogy, D., 2001, “Modifications to the DIRECT Algorithm”, Technical Report 01-013, Computer Mechanics Laboratory, University of California at Berkeley.

5. Zhu, H. and Bogy, D., 2001, “The CML Air Bearing Optimization Program Version 2.0”, Technical Report 01-016, Computer Mechanics Laboratory, University of California at Berkeley.

6. Zhu, H. and Bogy, D., 2002, "DIRECT Algorithm and Its Application To Slider Air Bearing Surface Optimization", *Intermag 2002 Conference*, Amsterdam, The Netherlands.

7. Zhu, H. and Bogy, D., 2002, "Hard disk drive air bearing design: Modified DIRECT algorithm and its application to slider Air Bearing Surface optimization", abstract accepted by *6th International Tribology Conference (AUSTRIB '02)*, Western Australia.

8. Bogy, D., Wu, L., Zeng, Q-H and Zhu, H., 1999, "Air Bearing Designs for 100 Gbit/in^2", TRIB-Vol. 9, *Proceedings of the Symposium on Interface Technology Towards 100 Gbit/in^2*, ASME, pp. 11-16

9. Bogy, D., Fong, W., Thornton, B., Zhu, H., Gross, H. and Bhatia, C., 2002, "Some Tribology and Mechanics Issues for 100 Gbit/in^2 HDD", *Intermag 2002 Conference*, Amsterdam, The Netherlands.

10. Ingber, L., 1989, "Very Fast Simulated Re-Annealing", *J Math. Comput. Modeling*, Vol. 12, pp. 967-973.

11. Ingber, L., 1993, "Simulated annealing: Practice versus theory", *J Math. Comput. Modeling*, 18, 11, pp. 29-57.

12. Ingber, L., 1996, “Adaptive Simulated Annealing (ASA): Lessons learned”, *Journal of Control and Cybernetics*, Vol. 25, pp. 33-54.

13. Jones, D. R., Perttunen, C. D. and Stuckman, B. E., 1993, “Lipschitzian Optimization Without the Lipschitz Constant ”, *Journal of Optimization Theory and Application*, Vol. 79, No. 1, pp. 157-181.

14. Gablonsky, J. M., 1998, “An Implementation of the DIRECT algorithm”, Technical Report CRSC-TR98-29, Center for Research in Scientific Computation, North Carolina State University.

15. Gablonsky, J. M. and Kelley, C. T., 2000, “A locally-biased form of the DIRECT algorithm”, Technical Report CRSC-TR00-31, Center for Research in Scientific Computation, North Carolina State University.

16. Gablonsky, J. M., 2001, “DIRECT Version 2.0 User Guide”, Technical Report CRSC-TR01-08, Center for Research in Scientific Computation, North Carolina State University.

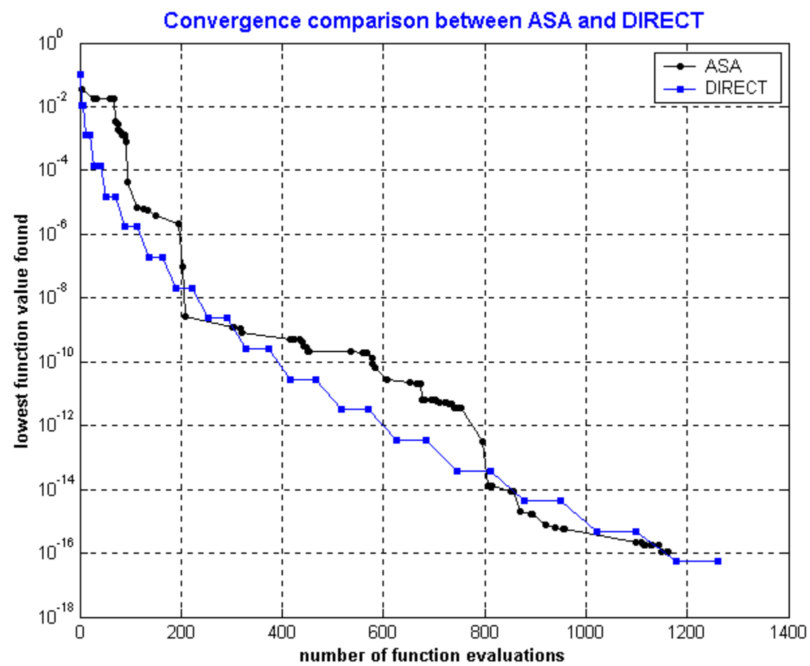


Fig. 1 Convergence comparison for 2-D case

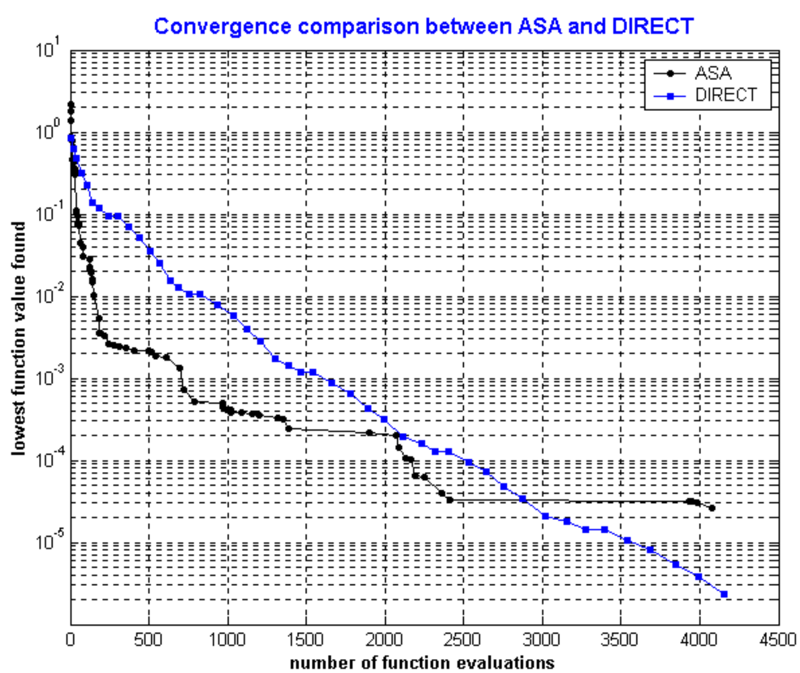


Fig. 2 Convergence comparison for 10-D case

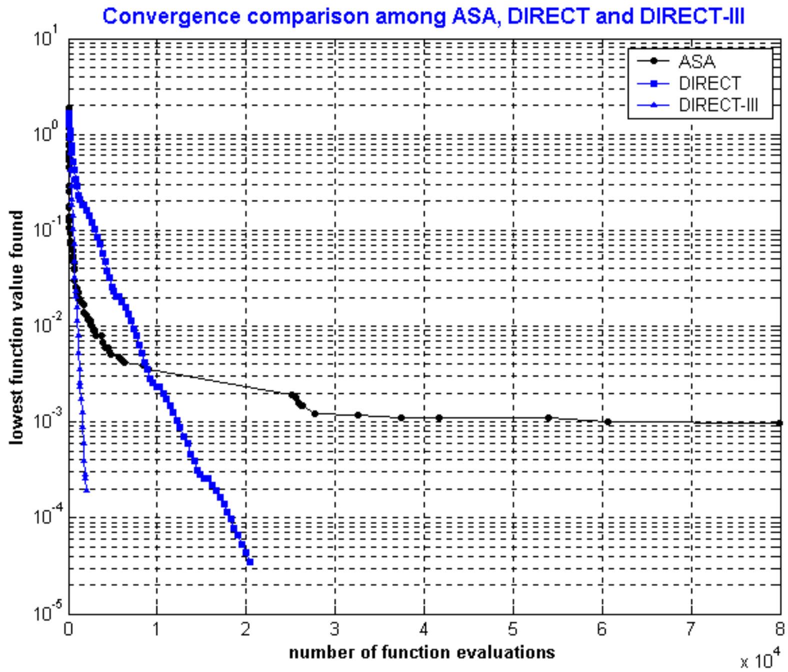
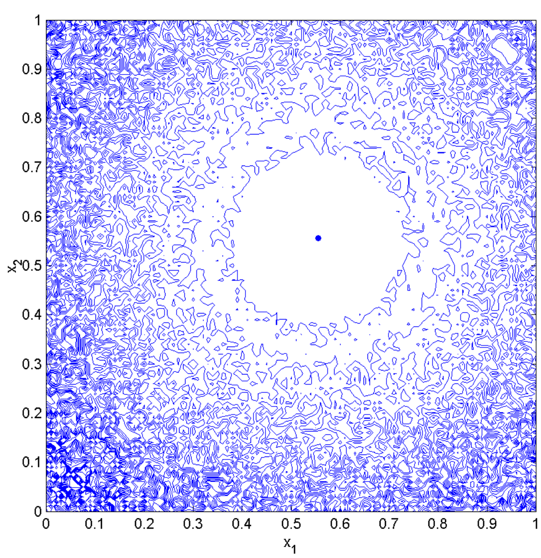


Fig. 3 Convergence comparison for 20-D case



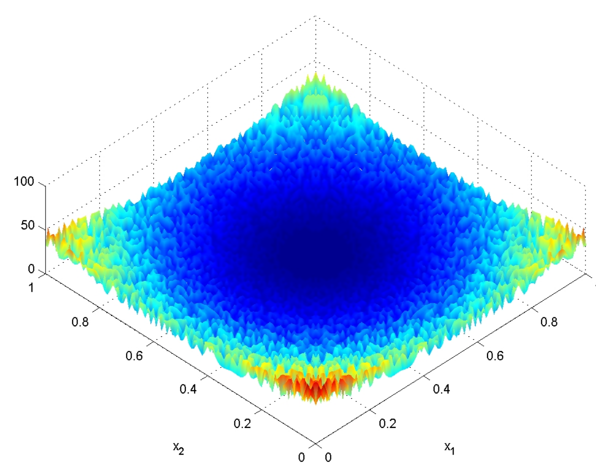


Fig. 4 Contour lines

Fig. 5 Surface shape

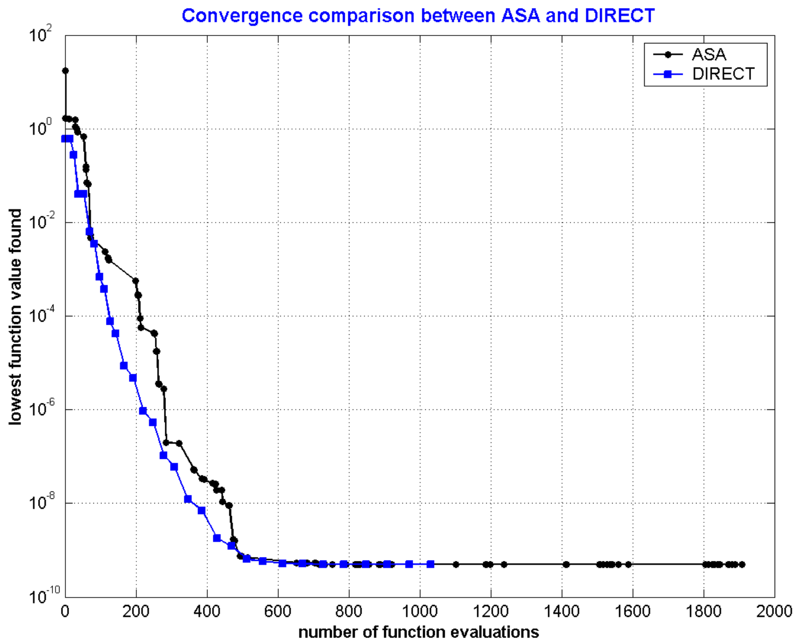


Fig. 6 Convergence comparison

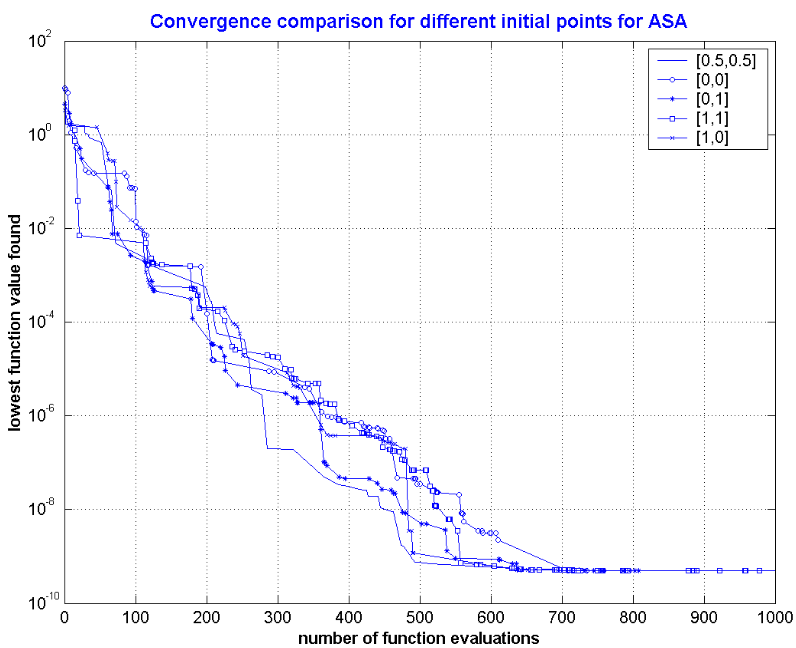


Fig. 7 Convergence comparison for different initial points using ASA

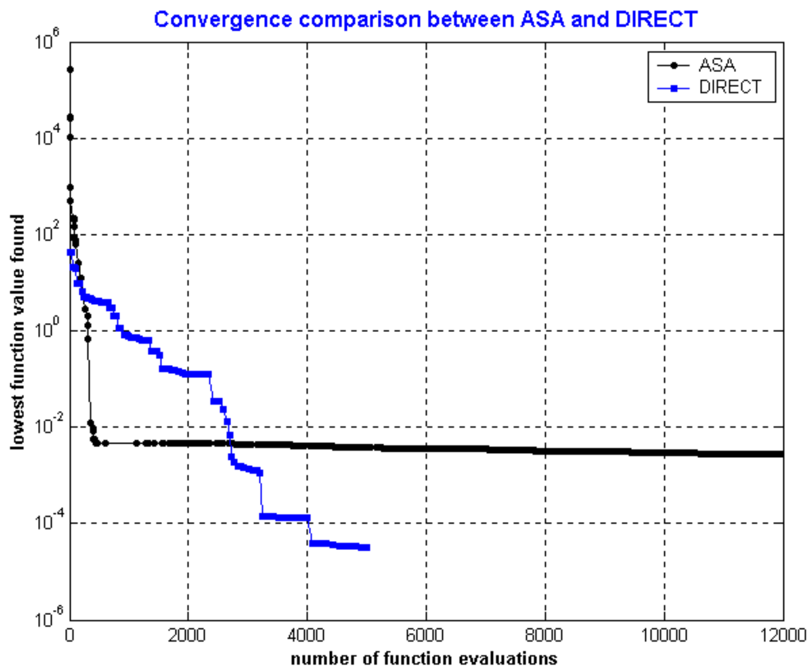
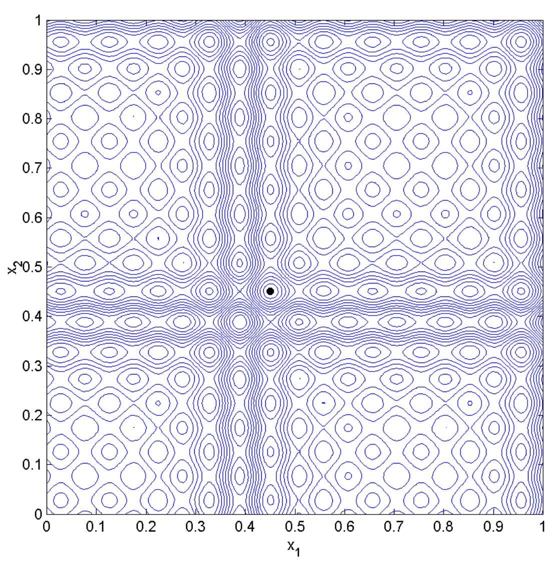


Fig. 8 Convergence comparison for 4-D Colville function



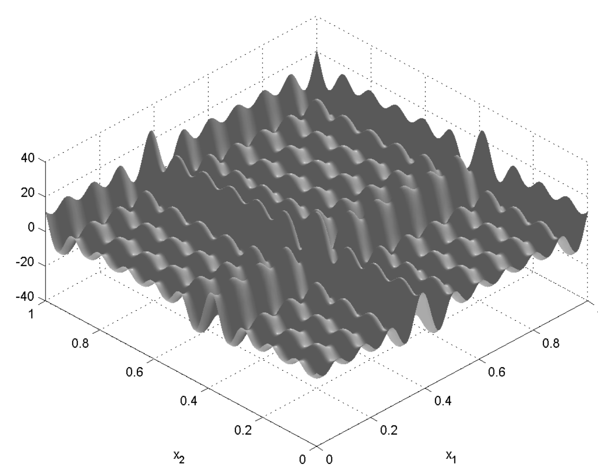


Fig. 9 Contour lines

Fig. 10 Surface shape

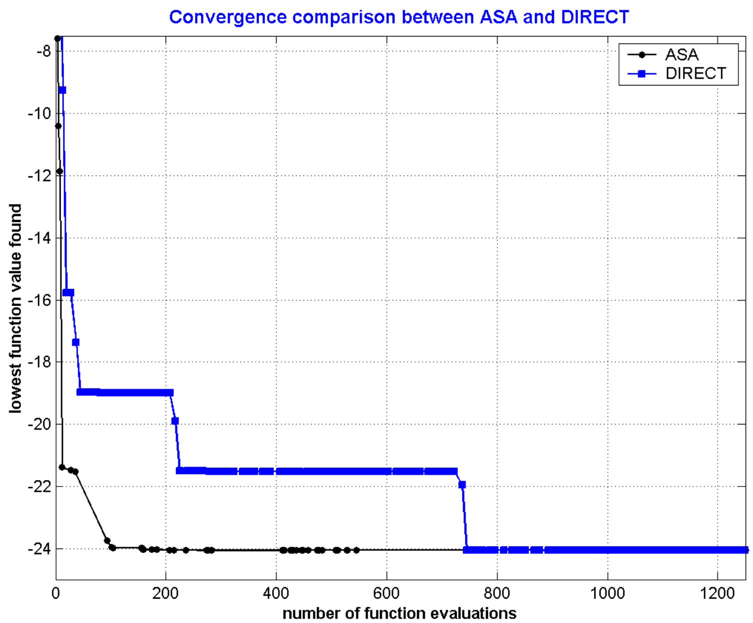


Fig. 11 Convergence comparison for local Shubert function

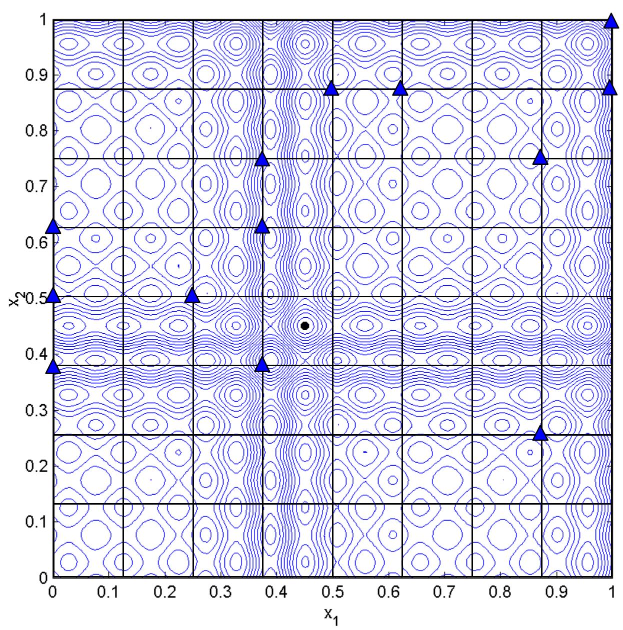


Fig. 12 Optimization results for different initial points for ASA

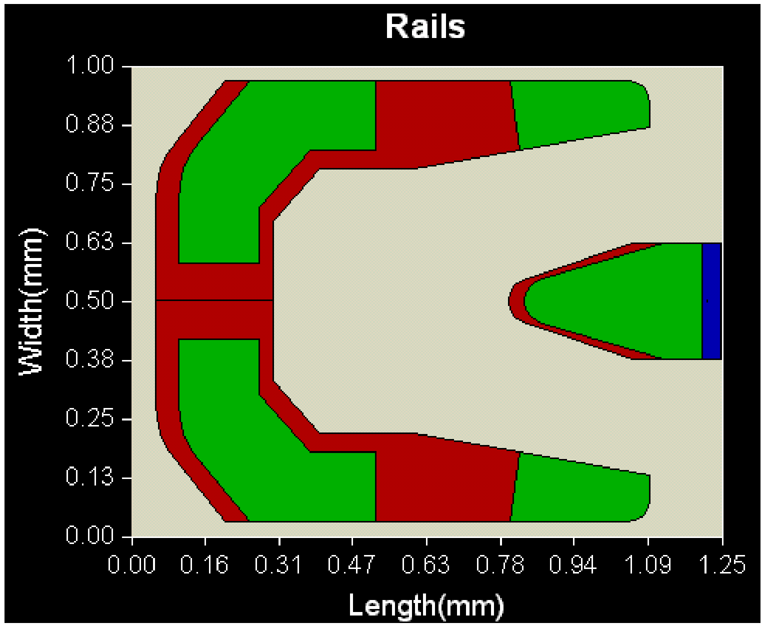


Fig. 13 Rail shape of the initial ABS design

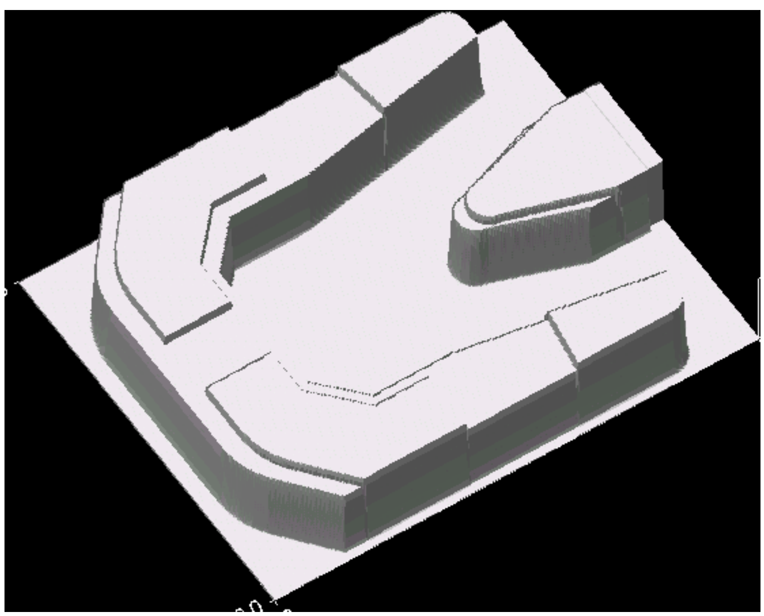


Fig. 14 3-D rail shape of the initial ABS design

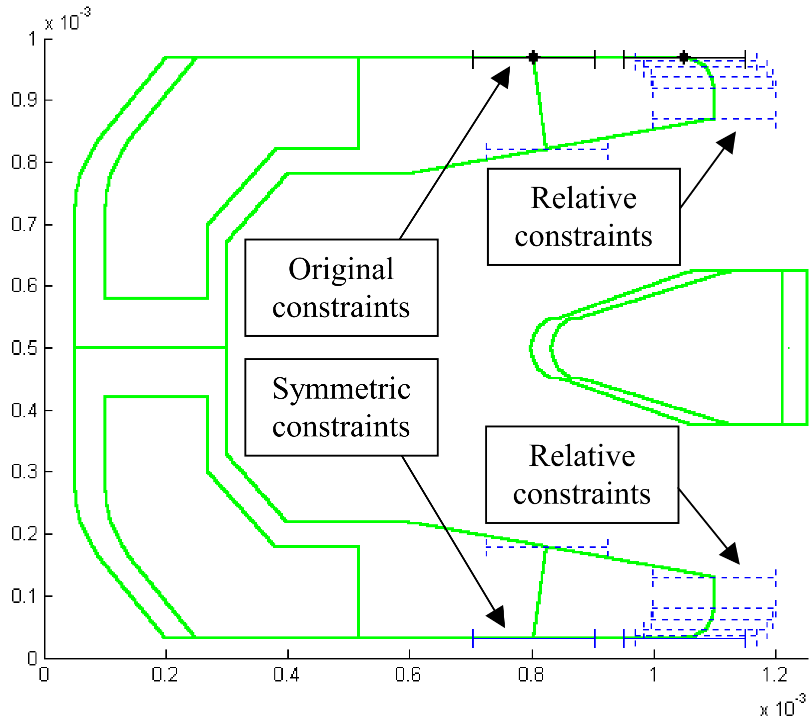


Fig. 15 Constraints defined on the initial design in the 2-D case

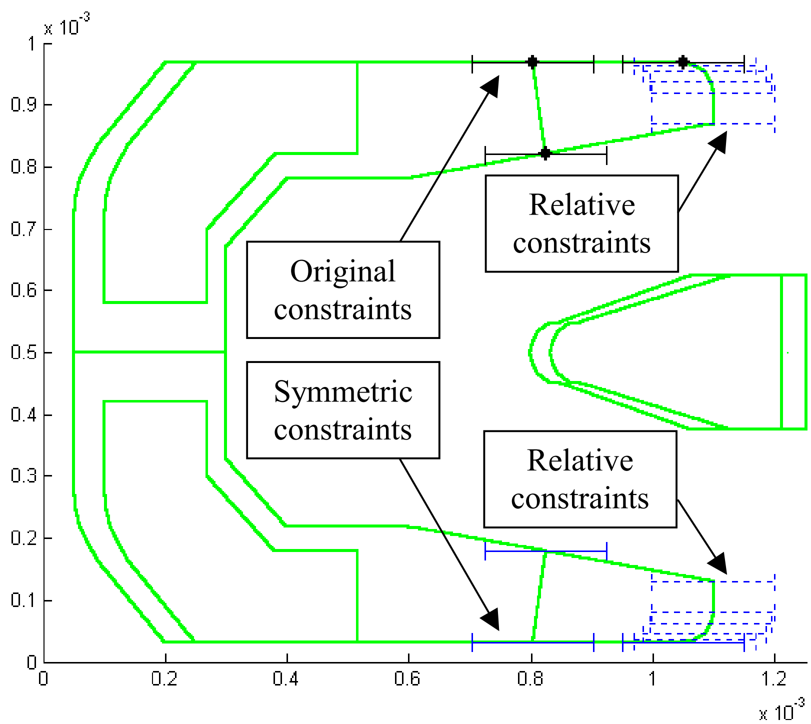


Fig. 16 Constraints defined on the initial design in the 3-D case

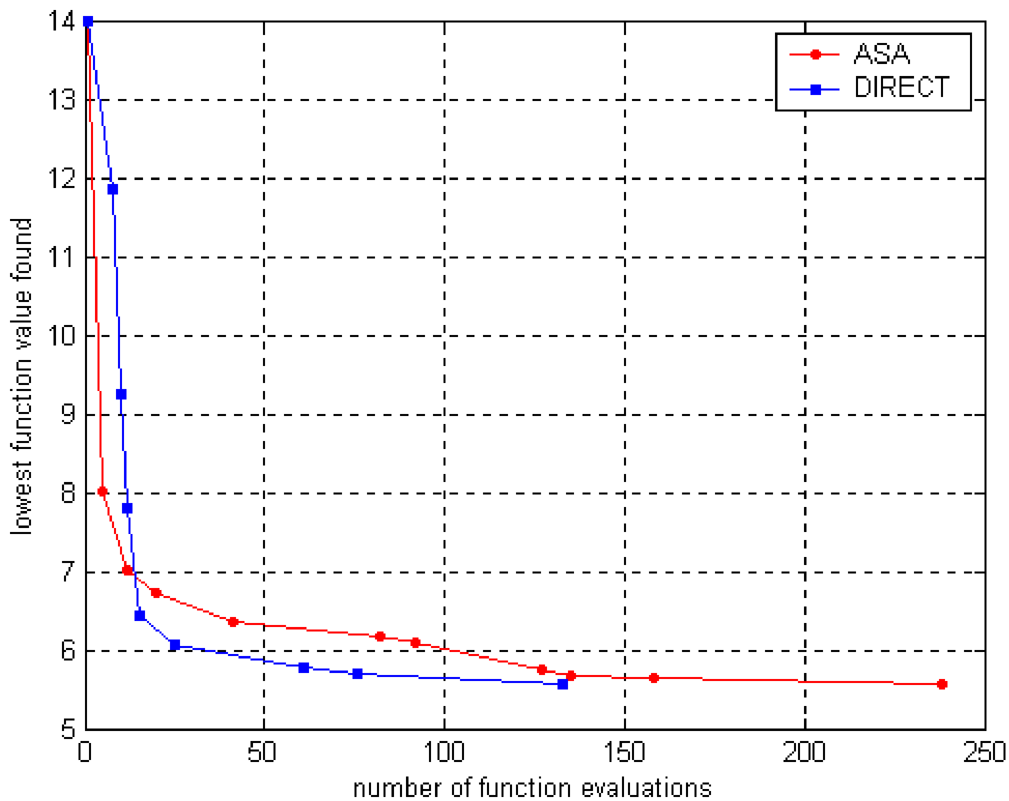


Fig. 17 Convergence comparison for the 2-D ABS optimization problem

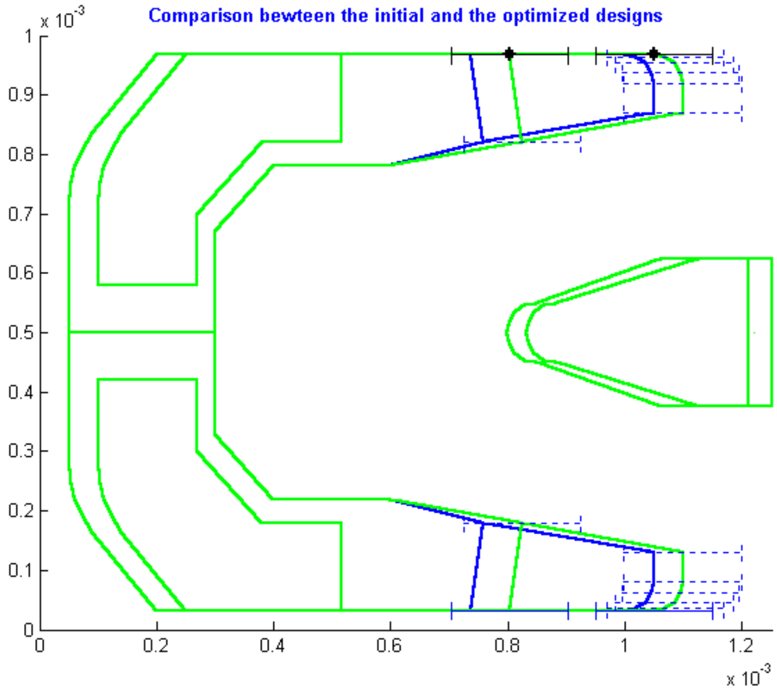


Fig. 18 Optimized design obtained by DIRECT in the 2-D case

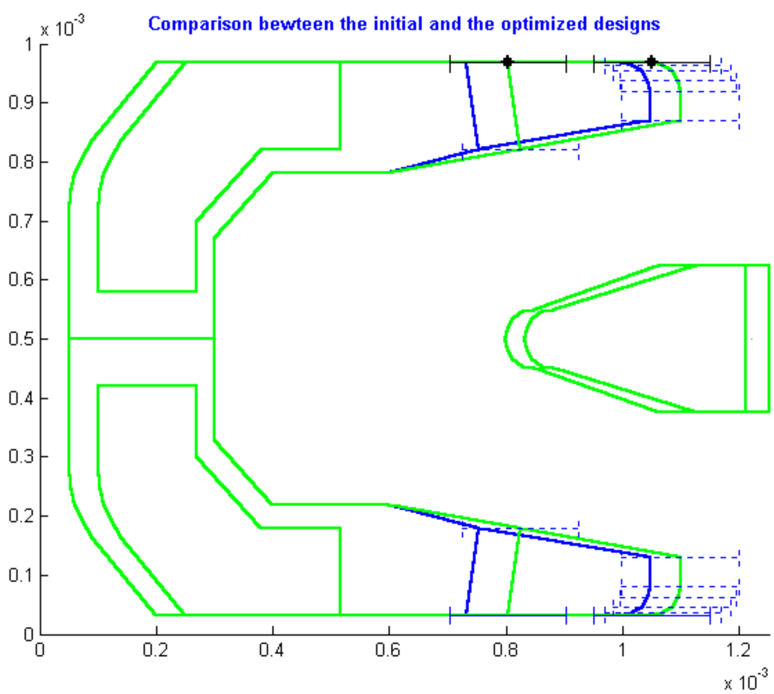


Fig. 19 Optimized design obtained by ASA in the 2-D case

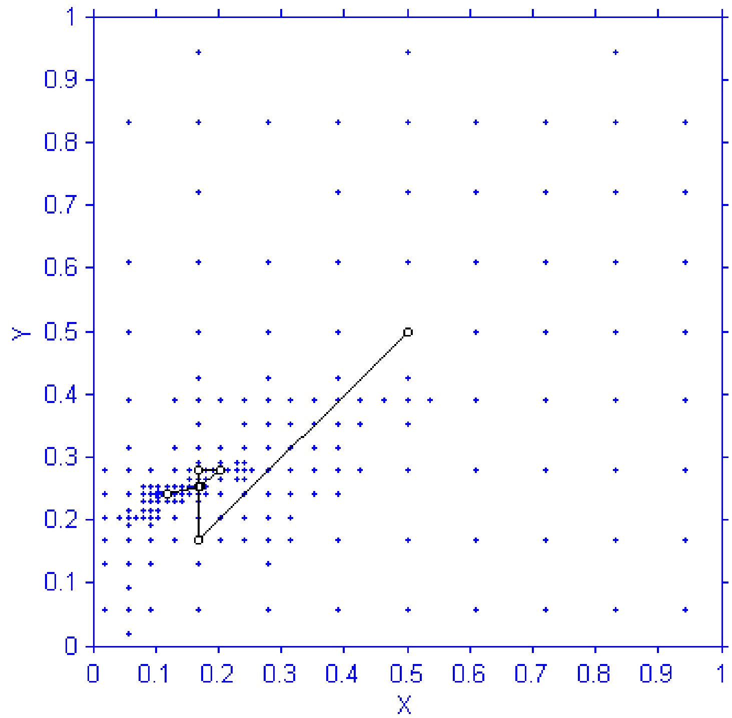


Fig. 20 Optimization results using DIRECT in the 2-D case

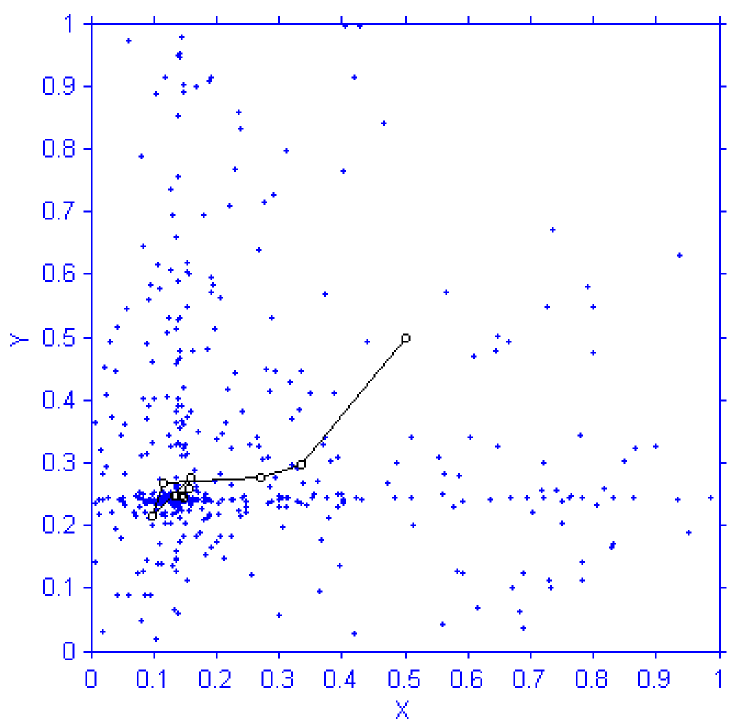


Fig. 21 Optimization results using ASA in the 2-D case

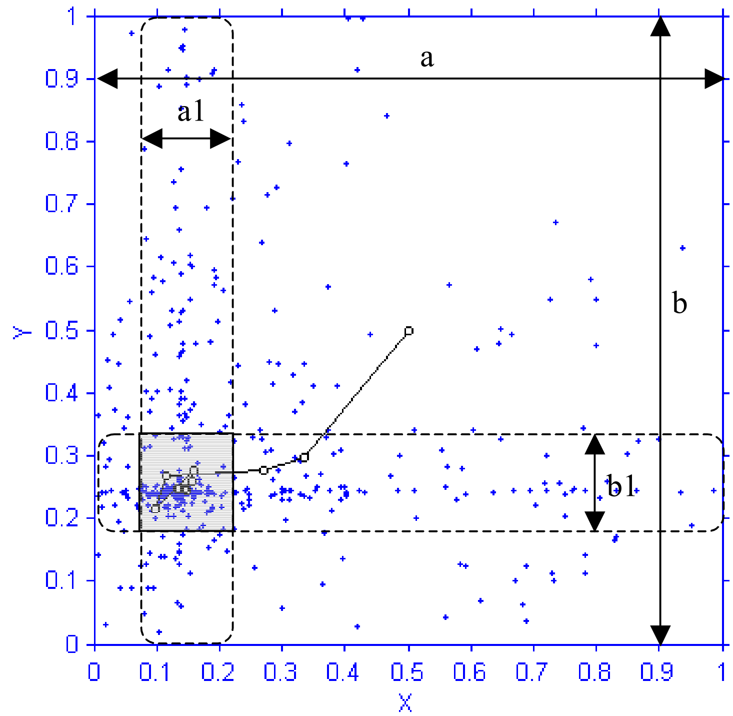


Fig. 22 Illustration of the “band” pattern of the optimization results using ASA

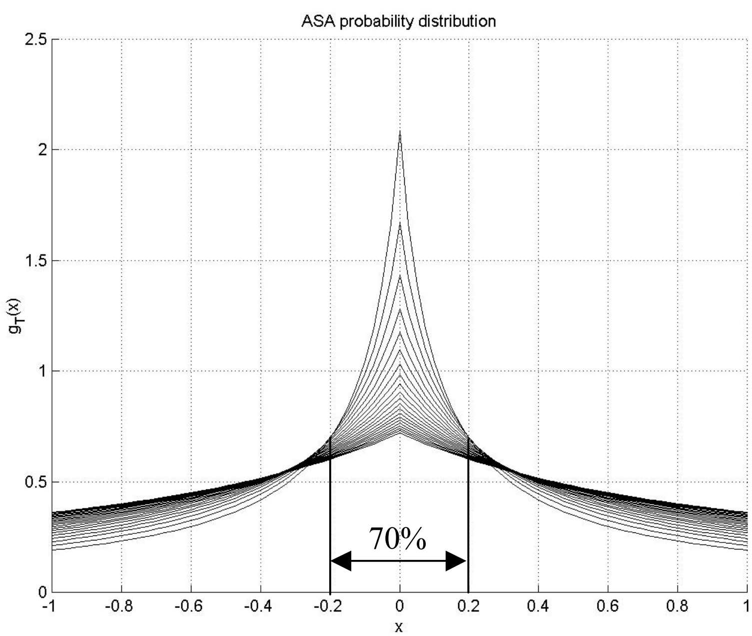


Fig. 23 ASA probability function

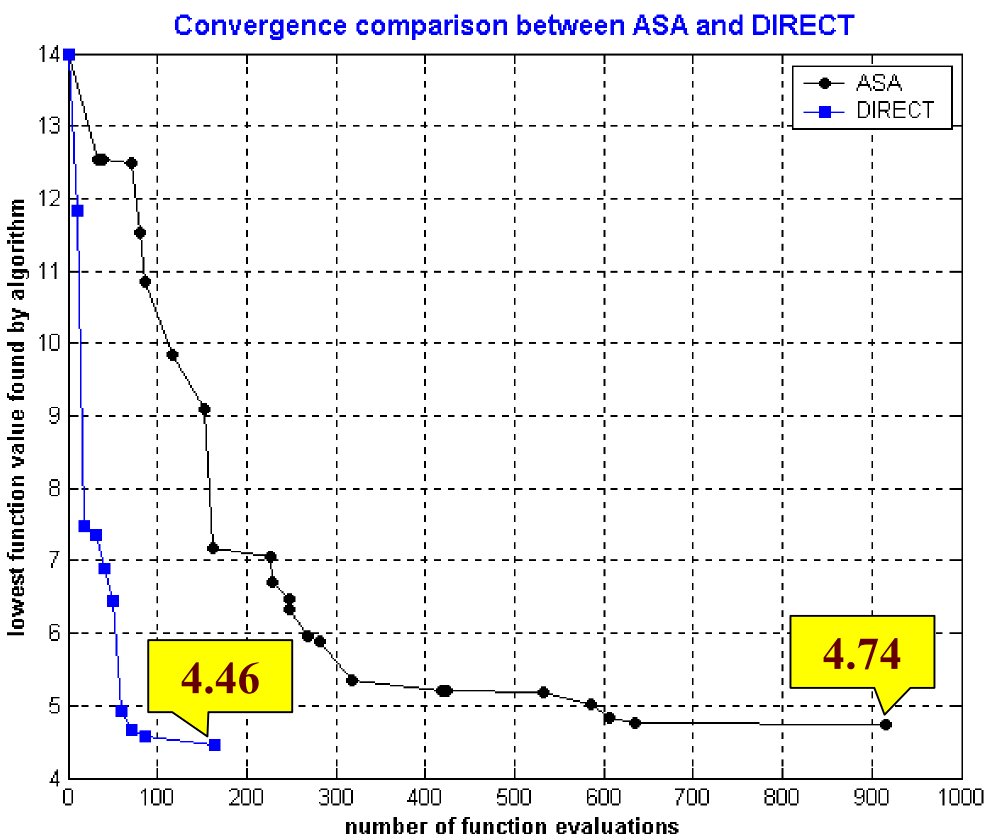


Fig. 24 Convergence comparison for the 3-D ABS optimization problem

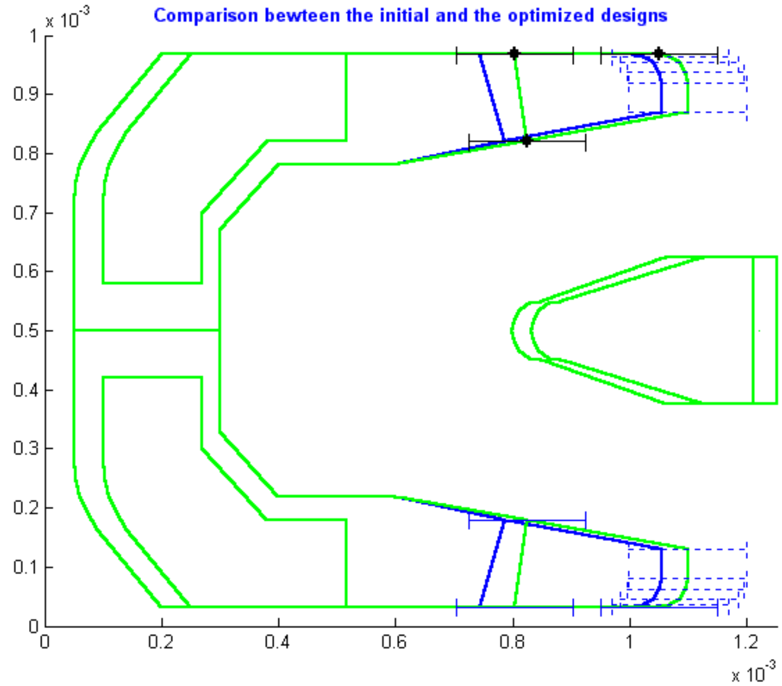


Fig. 25 Optimized design obtained by DIRECT in the 3-D case

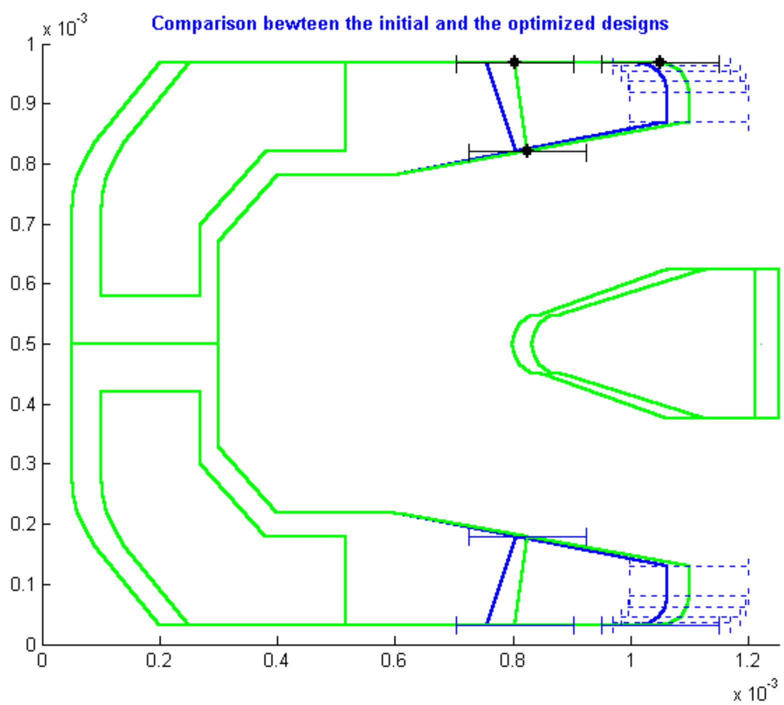


Fig. 26 Optimized design obtained by ASA in the 3-D case

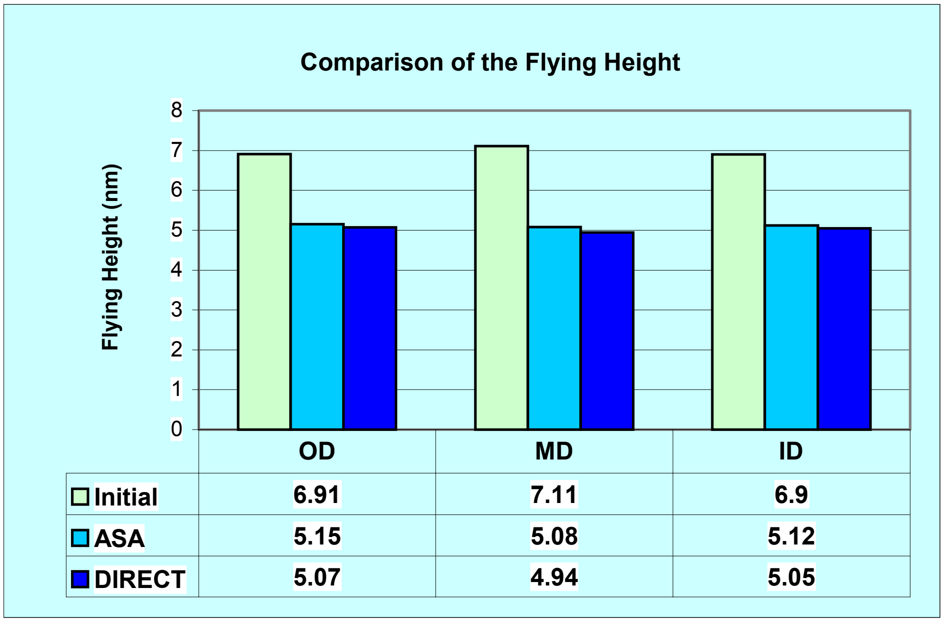


Fig. 27 Comparison of the FH for the 3-D ABS optimization case

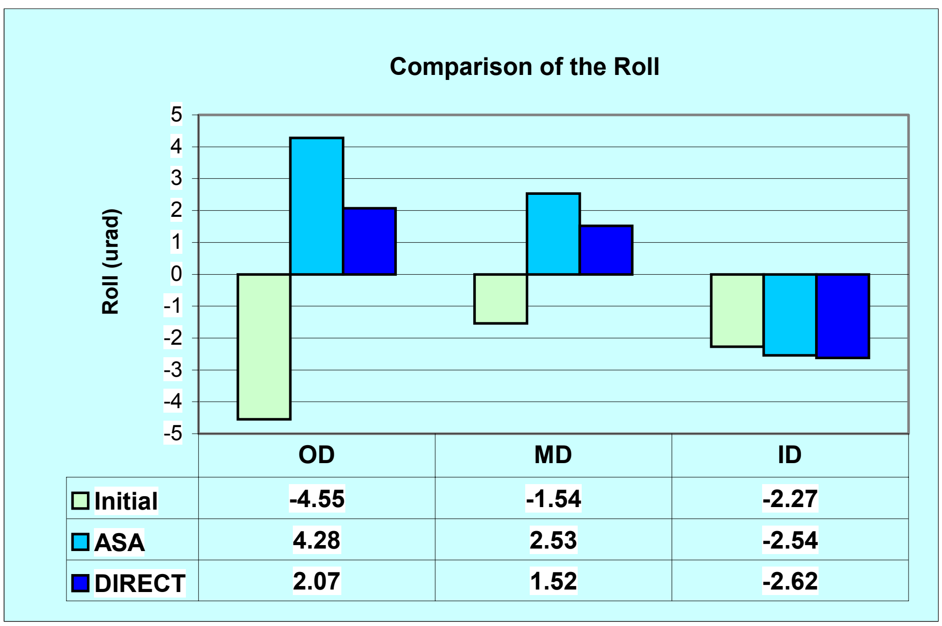


Fig. 28 Comparison of the Roll for the 3-D ABS optimization case