Multimodality and Global Optimization in Aerodynamic Design

Oleg Chernukhin and David W. Zingg

Two optimization algorithms are presented that are capable of finding a global optimum in a computationally efficient manner: a gradient-based multistart algorithm based on Sobol sampling and a hybrid optimizer combining a genetic algorithm with a gradient-based algorithm. The optimizers are used to investigate multimodality in aerodynamic-shape-optimization problems. The performance of each algorithm is tested on an analytical test function as well as several aerodynamic-shape-optimization problems in two and three dimensions. In each problem the primary objectives are to classify the problem according to the degree of multimodality and to identify the preferred optimization algorithm for the problem. The results show that multimodality should not always be assumed in aerodynamic-shape-optimization problems. Typical two-dimensional airfoil-optimization problems are unimodal. Three-dimensional shape-optimization problems may contain local optima. The number of local optima tends to increase with increasing geometric degrees of freedom and design space bounds. For problems with a modest number of local optima, which we term somewhat multimodal, the gradient-based multistart Sobol algorithm is the most efficient method.

I. Introduction

The use of computer algorithms for aerodynamic shape optimization (ASO) has the potential to uncover unconventional aircraft configurations that can lead to dramatic reductions in drag. This may result in significant improvements in aircraft fuel efficiency, which is important because of rising fuel prices and concerns over the environmental impact of aviation.

The two major components of ASO are efficient computational-fluid-dynamics (CFD) solvers and optimization algorithms. Although CFD has become a mature technology and has found numerous industrial applications, the use of optimization for CFD-based design is still undergoing rapid development.

The field of optimization is expansive, and the choice of a suitable algorithm is highly problem dependent. Considerations must be made with regard to the types of design variables (e.g., discrete and/or continuous), the number of constraints, the properties of the design space (e.g., number of local optima, discontinuities), etc. In this work we assume that the design space is smooth, and the design variables are continuous. The smoothness of the design space is a reasonable assumption if the numerical error is small and tight convergence criteria are specified. The authors’ experience suggests that, for a wide range of problems, these assumptions are valid.

Even with these assumptions many choices of optimization algorithms are available for ASO [1–3]. Traditionally, optimization algorithms have been divided in two broad categories: gradient-free and gradient-based (GB) methods. Both types have been used for ASO with compelling arguments presented in favour of each method. Moreover, various hybrid approaches incorporating elements from both GB and gradient-free algorithms have been proposed and successfully applied [3–9]. These hybrid algorithms attempt to address the shortcomings of the traditional optimization methods.

Gradient-based algorithms require sensitivities of the objective and the constraints in order to reach a local optimum. Quasi-Newton methods, such as Broyden-Fletcher-Goldfarb-Shanno (BFGS), construct a Hessian approximation at each iteration. The main advantage of gradient-based algorithms is their rapid convergence. The difficulties associated with GB optimizers arise in the need for efficient gradient calculations and in their inherent tendency to converge to local optima.

Gradient-free approaches usually mimic some real-life phenomenon in an attempt to minimize the objective function. Because gradient information is not required, these algorithms can be easily incorporated into existing frameworks and have found many practical applications. Genetic algorithms (GAs) are among the most popular gradient-free methods in use today [2,10,11]. Aside from their ease of implementation GAs are particularly suitable for problems with discrete design variables, problems with discontinuous objectives, and problems with multiple local optima (multimodality). The main disadvantage of gradient-free optimizers is their slow convergence. Zingg et al. [12] show that for two-dimensional (2-D) airfoil-optimization problems a GA can require 200 times more function evaluations than an efficient GB method.

The issue of multimodality in high-fidelity ASO based on the Euler and Navier–Stokes equations is largely unsettled. To the authors’ knowledge no definitive investigation of multimodality in high-fidelity ASO problems has been published to date. Some existing studies present results that appear multimodal as a result of insufficient optimizer convergence. For example, Namgoong et al. [13] conclude that because their gradient-based algorithm generates different results depending on the initial condition, the problem is multimodal. However, this can also occur if the optimizations are not fully converged. Various additional publications also claim that multimodality makes gradient-based algorithms a poor choice for ASO but without providing evidence of multimodality [3]. The existence of multiple local optima may be challenging to disprove. However, it is possible to prove that a local optimum has been reached, provided that gradient information is available.

Unfortunately, proof of optimization convergence is often omitted in ASO literature.

†Presented as Paper 2011-3070 at the 20th AIAA Computational Fluid Dynamics Conference, Honolulu, HI, 27–30 June 2011; received 5 January 2012; revision received 29 October 2012; accepted for publication 13 November 2012; published online 10 April 2013. Copyright © 2013 by Oleg Chernukhin and David W. Zingg. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. Copies of this paper may be made for personal or internal use, on condition that the copier pay the $10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 1533-385X/13 and $10.00 in correspondence with the CCC.

‡Former MSc Student, Institute for Aerospace Studies; currently Mathematics Software Developer, ANSYS Canada; olg.chemukhin@ansys.com.

§Professor and Director, Institute for Aerospace Studies, Tier 1 Canada Research Chair in Computational Aerodynamics, J. Armand Bombardier Foundation Chair in Aerospace Flight, Associate Fellow AIAA.
Nevertheless, multiple local optima certainly exist in aerodynamic-shape-optimization problems. Buckley et al. [14] present a practical multipoint airfoil optimization case and show that at least two local optima exist for this problem. Leung and Zingg [15] show that a three-dimensional (3-D) optimization of an ONERA M6 wing produces at least two local optima, one for forward and one for backward sweep. Hicken and Zingg [16] show that for spanwise vertical shape optimization at least two local minima exist as well (winglet-up and winglet-down).

In this work we propose two novel optimization algorithms that are designed to be computationally efficient and to avoid converging to local optima. The algorithms employ features from a GA and a GB algorithm as well as an efficient sampling process to thoroughly explore the design space. We first examine the algorithms on a highly multimodal test function and proceed to apply them to practical optimization problems. In addition, we compare them with two popular optimization methods that can be placed at the opposite ends of the gradient-free/gradient-based spectrum. The objectives of this paper are to investigate the multimodality of the ASO design spaces and to introduce and assess two new algorithms for dealing with multimodal design spaces.

II. Overview of the Integrated Aerodynamic-Shape-Optimization Methods

The key components of the aerodynamic-shape-optimization algorithms used are thoroughly described by Nemec and Zingg [1,17] for 2-D problems and Hicken and Zingg [18,19] for 3-D problems. Some of the main features of these algorithms are briefly summarized next.

The geometry parameterization is accomplished using B-spline curves in 2-D and B-spline surfaces in 3-D. The design variables in 2-D are vertical coordinates of the B-spline control points. In 3-D the design variables are the x, y, and z coordinates of the B-spline surface control points. Depending on the problem definition each control point is assigned from zero to three design variables.

In the 2-D algorithm the algebraic mesh movement algorithm of Nemec and Zingg [1] is used. In the 3-D algorithm the linear elasticity method of Truong et al. [20] is employed but applied to a B-spline volume control-point mesh as described by Hicken and Zingg [19].

The governing equations in 2-D are the compressible Navier-Stokes equations with the Spalart-Allmaras turbulence model used to compute the eddy viscosity. The discretized equations are solved using a Newton-Krylov approach [1].

In 3-D the governing equations are the Euler equations. Although the work to incorporate viscous terms and turbulence modelling is underway [21], the current study considers only inviscid 3-D problems. The spatial discretization is accomplished using summation-by-parts operators, and the interface conditions between blocks are enforced using simultaneous approximation terms [18]. The discretized equations are solved using a Newton-Krylov-Schur approach [18].

In both cases derivatives of objectives and constraints with respect to the design variables are computed using a discrete adjoint method. The adjoint method is the most efficient method when the number of design variables exceeds the number of nonlinear constraints, which is the case for most aerodynamic-shape-optimization problems. The chief advantage of the adjoint method is that the gradient calculation is nearly independent of the number of design variables.

III. Optimization Components

In general terms an optimization problem can be stated as follows:

\[
\begin{align*}
\text{minimize } & \mathcal{J}_o(X), \\
\text{w.r.t. } & X, \\
\text{s.t. } & a \leq \mathcal{J}_m(X) \leq b, \quad j = 1, \ldots, n_m, \\
& c \leq \mathcal{J}_l(X) \leq d, \quad k = 1, \ldots, n_l
\end{align*}
\] (1)

In the preceding formulation \( \mathcal{J}_o \) is the aerodynamic objective function to be minimized, often \( C_D \) or \( C_L/C_D \), \( X \) is a vector of design variables, such as the coordinates of the B-spline control points and/or the angle of attack \( \alpha \), and \( \mathcal{J}_m \) and \( \mathcal{J}_l \) are nonlinear and linear constraints, respectively. Examples of these are a wing-volume constraint, a lift constraint, and upper/lower bounds on the design variables (box constraints).

Nonlinear constraints can be incorporated into the objective function using a quadratic penalty method (QPM):

\[
\mathcal{J}(X, \rho) = \mathcal{J}_o(X) + \rho \sum_{i=1}^{n_c} \{\max(0, a - \mathcal{J}_m(X))\}^2 \\
+ \rho \sum_{j=1}^{n_l} \{\max(0, \mathcal{J}_l(X) - b)\}^2
\] (2)

This method can be applied to linear constraints as well, but these can be satisfied exactly. The choice of the appropriate penalty parameter \( \rho \) is not a trivial task especially for practical design problems [14]. The QPM is often used with gradient-free optimizers. However, for gradient-based optimization algorithms the preferred way to satisfy constraints is through the solution of the Karush-Kuhn-Tucker equations using the sequential-quadratic-programming (SQP) methodology [22]. In this work, both QPM and SQP techniques are employed.

A. Sobol Sampling

Effective sampling is important for design-space exploration. Although random sampling can be employed, more efficient strategies have been devised. Some of the most popular sampling techniques include Latin hypercube sampling and Sobol sequences. In this work we chose the Sobol sequence due to its deterministic behaviour and the ability to perform incremental sampling. It was originally introduced with the goal of approximating an integral of a d-dimensional function on a unit hypercube with the fastest possible convergence. We use an extension of algorithm 659, which uses Gray code implementation for generating Sobol sequences [23]. To generate the \( d \)th dimension of the \( n \)th sample point \( x_{d,n} \) we use a recursive relation:

\[
x_{d,1} = 0 \quad \text{and} \quad x_{d,n} = x_{d,n-1} + \bigoplus v_{d,n-1}
\] (3)

where \( \bigoplus \) is a bitwise exclusive-or operator, \( c_0 \) is the index of the first 0 digit from the right of the binary representation of \( n \), and \( v_{d,n} \) is the directional number defined as:

\[
v_{d,n} = \frac{m_{d,n} - 2^c}{2^d}
\] (4)

A proper Sobol sequence requires a set of carefully generated directional numbers. The list of numbers \( m_{d,n} \) used in this work is provided by Joe and Kuo [23] and can be used for dimensions up to \( d = 21, 201 \).

B. Gradient-Based SNOPT Optimization Algorithm

The gradient-based optimizer used is the optimization package SNOPT. Developed by Gill et al. [24] SNOPT uses an SQP algorithm to find the solution to nonlinear optimization problems with general constraints. The Hessian of the Lagrangian function is approximated using the quasi-Newton BFGS method.

In this paper we use SNOPT terms "merit function," "optimality tolerance," and "feasibility tolerance." These are described in detail by Gill et al. [24]. The merit function is an augmented Lagrangian merit function, which is equal to the objective function when all constraints are satisfied. The optimality tolerance specifies how small the gradient of the Lagrangian function must be, normalized by the Lagrange multipliers. The feasibility tolerance specifies how small the violation of the nonlinear constraints must be, normalized by the values of the design variables.
C. Gradient-Based Multistart Algorithm Based on Sobol Sampling

Because a gradient-based optimization algorithm, such as SNOPT, uses one initial guess and converges to a local optimum only, it is logical to consider an algorithm that uses multiple starting points to initiate gradient-based optimization. The advantage of incremental sampling is particularly important here. If a Sobol sequence is employed the user can optimize the first $n$ initial points in this sequence using a GB algorithm. Once these points are optimized the user can then increment the sampling and optimize the next $n$ initial points. This process can be repeated until a user-defined termination criterion has been met (e.g., time constraints, limits on computational resources, or the user is convinced that the design space has been thoroughly explored). The parallel implementation is done using the Message Passing Interface library. We use GB-MS to denote this algorithm.

To avoid infeasible geometries we created a system of geometric linear constraints for each problem. This system limits the geometric deformation by controlling a limited number of input parameters. One example of these constraints is shown next. The detailed description of the entire system of linear constraints is presented by Chernukhin [25].

Consider a chordwise cross section of a wing that is parameterized with B-spline surfaces. We define the following: 1) $U$ is the control point on the upper surface, 2) $L$ is the corresponding control point at the leading edge, 3) $T$ is the corresponding control point at the trailing edge, 4) $U_z$ is the vertical coordinate of $U$ (define $L_z$, $T_z$ similarly), 5) $U_x$ is the streamwise coordinate of $U$ (define $L_x$, $T_x$ similarly), 6) $\gamma$ is the streamwise fraction of $U$ along the chord line:

$$\gamma = \frac{U_x - T_x}{L_x - T_x}$$

7) $h_0$ is the initial height of $U$ above the chord line:

$$h_0 = U_z - \gamma L_z - (1 - \gamma) T_z$$

8) $\beta$ is the input parameter (see next equation).

With the help of the input parameter $\beta$ we can constrain the bounds of the vertical coordinate $U_z$ to the fraction of its initial height $h_0$ above the chord as follows:

$$(1 - \beta) h_0 \leq U_z - \gamma L_z - (1 - \gamma) T_z \leq (1 + \beta) h_0$$

For example setting $\beta = 0.6$ allows $U_z$ to vary between $0.4 h_0$ to $1.6 h_0$.

A similar process can be followed to define appropriate constraints for the remaining design variables. In the end every linear geometric constraint is expressed in the following form:

$$l \leq a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \leq u$$

where $x_i$ are design variables, and $l, u, a_i$ are constants.

The advantage of linear constraints is that their gradients are readily known, which allows the optimizer to satisfy these constraints exactly. Using this method we can start with a generic geometry, shown in Fig. 1a, and set the linear constraints based on this geometry. We can then create an arbitrary sample, shown in Fig. 1b, and shape it into a wing by enforcing the linear constraints (Fig. 1c).

This process takes place in less than a second.

This method is an alternative to the approach of Leung and Zingg [15], where problem-specific design variables such as sweep are created. If custom design variables are introduced the section of the solver source code needs modifications to calculate sensitivities. In the current approach the constraints are problem specific, but the design variables are always of the same type (coordinates of the control points).

One disadvantage of creating arbitrary samples (Fig. 1b), is that after the linear constraints are enforced the sample is biased towards the boundaries of the linear-feasible region $R_L$. Figure 2 provides an illustration of this process. In Fig. 2a a 2-D sample is shown with the linear-feasible region in dashed lines. In Fig. 2b is the same sample with linear constraints enforced by SNOPT. Clearly, the majority of points are on the boundaries of $R_L$, and the interior of the region is not thoroughly sampled.

To address this issue we have designed a procedure for sampling only within $R_L$. The main idea is to provide the proper order in which the linear constraints are implemented. Then, it is straightforward to establish the upper and lower bounds on each design variable. Figure 2c shows a sample that results from using this procedure to sample only within $R_L$. Clearly, the linear-feasible region is now covered in a uniform, unbiased manner, an obvious improvement over the sample in Fig. 2b.

---

**Fig. 1** Wing parameterization.

**Fig. 2** Sampling method.
D. Genetic Algorithm

Genetic Algorithms (GA) have been used extensively in the aerospace field [2,11,12]. GAs attempt to minimize the objective function by mimicking the process of evolution. In the GA context the design points are called chromosomes, the objective function is the fitness, and genes refer to either the design variables or substrings of bit-encoded design-variable strings.

With numerous variations the basic operators of any GA are selection, crossover, and mutation. Once each chromosome is assigned a fitness value the selection process decides which chromosomes are considered in the creation of a new generation. The crossover operator combines various genes from the two chromosomes (parents) to create a new one (child). The mutation operator assigns random values to some of the genes in a chromosome. Through the process of selection, crossover, and mutation, the subsequent generations are created. With the aid of the CFD solver the fitness values are assigned to the chromosomes in the new generation, and the process continues.

The GA used in this project was developed at the NASA Ames Research Center by Holst and Pulliam [26]. It uses real number encoding, which is more practical when all design variables are continuous. In order to create the next generation this GA uses four basic operators: passthrough, crossover, pure mutation, and perturbation mutation. The input parameter \( p \)-vector states what percentage of the chromosomes within the new generation are created using a particular operator. For example, \( p = [0.1, 0.2, 0.3, 0.4] \) states that 10% of the chromosomes are created using passthrough, 20% using crossover, 30% using pure mutation, and 40% using perturbation mutation.

The passthrough operator takes the fittest chromosomes from the previous generation and passes them on to the next generation. This ensures that information from the fittest chromosomes is never lost. The crossover operator combines two chromosomes in the following way:

\[
x_{\text{new}} = 1/2(x_1 + x_2)
\]

(9)

The pure-mutation operator takes one chromosome from the previous generation and replaces some of its genes with random values:

\[
x_{\text{new}} = \text{RAND}(x_{\text{max}} - x_{\text{min}}) + x_{\text{min}}
\]

(10)

where RAND is a random number between 0 and 1. The perturbation mutation takes one chromosome from the previous generation and perturbs some of its genes:

\[
x_{\text{new}} = x_{\text{old}} + \beta (x_{\text{max}} - x_{\text{min}})(\text{RAND} - 0.5)
\]

(11)

where \( \beta \) is a user-specified parameter. Checks are in place to ensure that \( x_{\text{new}} \) does not exceed its box constraints. For the GA optimizer the \( p \)-vector used in this work is:

\[
p = \left[ \frac{1}{n} \frac{1}{2} \frac{1}{n} \frac{1}{4} \frac{1}{4} \right]
\]

(12)

where \( n \) is the population size. Nonlinear constraints are enforced using the QPEN with a penalty term weight \( \rho = 10 \). The linear constraints are satisfied exactly using SNOPT, as described in the preceding section.

E. Hybrid Optimization Algorithm

Numerous ideas for hybrid optimization algorithms (HM) can be found in the literature [3,4,6–9,27,28]. Although many algorithms involve use of surrogate modeling to interpolate the objective function, this approach is not taken in this paper because surrogate models can introduce difficulties when the number of design variables is large.

We propose a hybrid optimizer that takes advantage of the GA’s ability to perform a global search and SNOPT’s ability to efficiently find the nearest local optimum and enforce constraints. Whenever a new generation is created by the GA the resulting chromosomes are passed to SNOPT for gradient-based refinement. We limit the number of SNOPT major iterations to equalize the amount of time SNOPT spends on each chromosome, which improves load balancing. However, care must be taken not to limit gradient-based iterations excessively within a single generation. SNOPT typically requires a problem-dependent minimum number of iterations that must be performed continuously (i.e., within a single generation) to reach optimality.

The improved chromosomes are then passed back to the GA to create a new generation. The block diagram of the hybrid optimizer is shown in Fig. 3.

The perturbation-mutation operator is designed to improve the GA refinement capability. Because we use a gradient-based optimizer for refinement, the perturbation mutation is avoided in the hybrid optimizer. For the HM optimizer the \( p \)-vector is:

\[
p = \left[ \frac{1}{n} \frac{1}{2} \frac{1}{n} \frac{1}{4} \frac{1}{4} \right]
\]

(13)

IV. Optimization Problems

The difficulty with assessing the performance of any optimization algorithm is that one algorithm may be more suited than other, depending on the optimization problem. Our main motivation in the development of the optimization algorithms is the ability to handle design spaces with multiple local optima. We classify optimization problems according to the number of local optima as shown in Table 1.

Unless otherwise stated each optimization problem is solved using the four optimization methods described in the preceding section: 1) Gradient-based method, single (baseline) initial guess (GB); 2) Gradient-based method, multiple initial guesses determined by Sobol sequence (GB-MS); 3) Hybrid optimization method (HM); and 4) GA with linear constraints handled by SNOPT (GA). Every problem will be classified according to the criteria in Table 1, and the most suitable optimization method for this problem will be selected. We assess the efficiency of these methods based on how many

<table>
<thead>
<tr>
<th>Number of local optima ( l )</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l = 1 )</td>
<td>Unimodal</td>
</tr>
<tr>
<td>( 1 &lt; l \leq 10 )</td>
<td>Somewhat multimodal</td>
</tr>
<tr>
<td>( 10 &lt; l \leq 100 )</td>
<td>Moderately multimodal</td>
</tr>
<tr>
<td>( l &gt; 100 )</td>
<td>Highly multimodal</td>
</tr>
</tbody>
</table>

![Fig. 3 Hybrid-optimizer block diagram.](image-url)
function evaluations they take to reach the best optimal point. For methods that use gradient information (GB, GB-MS, and HM), we count a gradient computation as one function evaluation because the computational expense of one adjoint solution approximately equals that of one flow solution. Thus, one objective evaluation and one gradient evaluation count as two function evaluations.

A. Optimization of a Highly Multimodal Analytical Test Function

Before ASO problems are considered the performance of the algorithms is assessed on an analytical test function. The Griewank function is particularly suitable for testing of the global optimization algorithms because it is smooth, differentiable, and highly multimodal. Its definition in two dimensions is given next:

$$f(x_1, x_2) = \frac{x_1^2}{4000} + \frac{x_2^2}{4000} - \cos \left( \frac{x_1}{\sqrt{2}} \right) \cos \frac{x_2}{\sqrt{2}} + 1$$  \hspace{1cm} (14)$$

This function has a global optimum $f(0, 0) = 0$. Its main features can be understood by examining Fig. 4 where the Griewank function is plotted over different domains. Although the function looks convex when plotted over an expansive domain, its highly multimodal nature becomes clear when a smaller region of the domain is examined.

For the highly multimodal test case the box constraints are:

$$-1234 \leq x_1 \leq 1321, \quad -934 \leq x_2 \leq 778$$

These constraints are arbitrary the only requirement being that the domain must not be centered on the origin (global minimum) because the second point in the Sobol sequence is placed at the exact center of the domain. Clearly, this function contains numerous local optima and the single-initial-guess GB optimizer cannot be used for this problem. Due to the stochastic nature of the GA and HM optimizers 10 optimization runs are performed using these algorithms, and the geometric mean of these 10 runs is calculated and plotted. Holst and Pulliam [26] perform a rigorous investigation of the number of GA runs required to obtain a true average of the convergence process. The results show that a reasonable average (well within 5% of the asymptote) is obtained after 10 runs for the convergence tolerance of $10^{-5}$. Although increasing multimodality and a smaller convergence tolerance may increase the variation in the GA runs, we chose 10 as a reasonable number to assess the performance of the GA. Because the multistart (GB-MS) procedure is fully deterministic (a property of the Sobol sequence), only one run is required. The population size is 16 for both the GA and the HM. Our preliminary results confirmed the conclusion of Holst and Pulliam [26] that the effect of population size on convergence efficiency is small.

We consider the total number of function evaluations it takes to reach the global minimum to within $10^{-8}$. The convergence plots are shown in Fig. 5a. For the GA and the HM the chromosome with the best fitness value is plotted at the end of each generation. For the GB-MS procedure we continuously sample using a Sobol sequence. After each sample point is driven to optimality we plot the best objective value found.

Of the three algorithms considered the GA consistently requires the largest number of function evaluations to converge. The GB-MS method required 27,643 function evaluations and 1907 Sobol sample points to find the global minimum. The hybrid optimizer significantly outperforms the other algorithms for this highly multimodal problem.

B. Optimization of a Moderately Multimodal Analytical Test Function

As in the previous problem we consider the Griewank test function 14. For the moderately multimodal problem the box constraints are:

$$-12.34 \leq x_1 \leq 3.76, \quad -11.41 \leq x_2 \leq 8.45$$
As in the previous problem, these constraints are arbitrary with care taken not to position the global minimum at the exact center of the domain. There are 18 local optima in this case. One can expect the advantage of the hybrid algorithm over the GB-MS method to be diminished because far fewer sample points are required to find the global minimum. As in the previous problem, we perform 10 optimization runs for each of the GA and HM algorithms, and one run for the GB-MS procedure.

The results support the hypothesis: the GB-MS and HM algorithms require a comparable number of function evaluations to reach the global optimum. This is evident from the convergence plots in Fig. 5b. The GA again shows the slowest convergence of the three algorithms tested.

Based on the results from these two test problems we can draw some intermediate conclusions. If the design space for a practical aerodynamic optimization problem is highly multimodal the HM algorithm may be the preferred choice. However, if the design space contains only a few local optima, the GB-MS optimizer may be most efficient in reaching a global optimum. With these thoughts we proceed to more practical optimization cases.

C. Airfoil-Optimization Problem

The 2-D ASO problem is solved using the Newton-Krylov algorithm described in Sec. II. The mesh is a 289 × 65 point C-mesh shown in Fig. 6. Length units are nondimensionalized, with chord length being equal to one. The geometry is parameterized using 23 B-spline control points. Two coincident control points at the trailing edge and one at the leading edge are kept constant. The remaining 20 control points are allowed to move in the vertical direction, subject to linear constraints as discussed in preceding sections. There are 10 control points on the upper surface and 10 on the lower surface. Their distribution is approximately uniform as determined by the fitting algorithm.

The optimization is performed at the following conditions:

$$Ma = 0.729, \quad Re = 7.0 \times 10^6, \quad \alpha = 2.31 \text{ deg}$$

where $Ma$ is the Mach number, $Re$ is the Reynolds number, and $\alpha$ is the angle of attack. We consider lift-constrained drag minimization with $C_l = 0.690$. The required lift coefficient is met by adjusting $\alpha$ as described by Billing [29]. A minimum area constraint of 0.07772 is also imposed.

For the GB optimizer the control points are located such that the initial geometry conforms to the standard RAE2822 airfoil. The initial CFD solution of this geometry is illustrated in Fig. 7a where the

![Fig. 6 C-grid for airfoil-optimization problem.](image)

As in the previous problem, these constraints are arbitrary with care taken not to position the global minimum at the exact center of the domain. There are 18 local optima in this case. One can expect the advantage of the hybrid algorithm over the GB-MS method to be diminished because far fewer sample points are required to find the global minimum. As in the previous problem, we perform 10 optimization runs for each of the GA and HM algorithms, and one run for the GB-MS procedure.

The results support the hypothesis: the GB-MS and HM algorithms require a comparable number of function evaluations to reach the global optimum. This is evident from the convergence plots in Fig. 5b. The GA again shows the slowest convergence of the three algorithms tested.

Based on the results from these two test problems we can draw some intermediate conclusions. If the design space for a practical aerodynamic optimization problem is highly multimodal the HM algorithm may be the preferred choice. However, if the design space contains only a few local optima, the GB-MS optimizer may be most efficient in reaching a global optimum. With these thoughts we proceed to more practical optimization cases.

C. Airfoil-Optimization Problem

The 2-D ASO problem is solved using the Newton-Krylov algorithm described in Sec. II. The mesh is a 289 × 65 point C-mesh shown in Fig. 6. Length units are nondimensionalized, with chord length being equal to one. The geometry is parameterized using 23 B-spline control points. Two coincident control points at the trailing edge and one at the leading edge are kept constant. The remaining 20 control points are allowed to move in the vertical direction, subject to linear constraints as discussed in preceding sections. There are 10 control points on the upper surface and 10 on the lower surface. Their distribution is approximately uniform as determined by the fitting algorithm.

The optimization is performed at the following conditions:

$$Ma = 0.729, \quad Re = 7.0 \times 10^6, \quad \alpha = 2.31 \text{ deg}$$

where $Ma$ is the Mach number, $Re$ is the Reynolds number, and $\alpha$ is the angle of attack. We consider lift-constrained drag minimization with $C_l = 0.690$. The required lift coefficient is met by adjusting $\alpha$ as described by Billing [29]. A minimum area constraint of 0.07772 is also imposed.

For the GB optimizer the control points are located such that the initial geometry conforms to the standard RAE2822 airfoil. The initial CFD solution of this geometry is illustrated in Fig. 7a where the

![Fig. 7 Comparison of the initial and optimized geometries in airfoil-optimization problem.](image)
Mach number contours are shown. The surface pressure coefficient distribution is displayed in Fig. 7c. The convergence plot for the GB optimizer, shown in Fig. 8a, demonstrates that a local optimum has been reached because the optimality criterion is reduced to $10^{-6}$. The optimized airfoil is shown in Fig. 8b. The Mach number contours of the optimized airfoil are shown in Fig. 7b, and the $C_p$ plot is provided in Fig. 7c. The drag coefficient is reduced from $1.518 \times 10^{-2}$ to $1.337 \times 10^{-2}$, a 12% improvement.

For GB-MS procedure we consider 480 initial guesses generated by the Sobol sequence. The first 10 initial guesses are shown in Fig. 8c. We use 100 generations for the HM optimizer and 1000 generations for the GA. None of the three global optimization algorithms were successful in finding additional local optima. The convergence plots are shown in Fig. 8d. Note that GB-MS and HM reach the same optimal point as the GB algorithm. However, the GA converges very slowly and is still not near the optimal point after 64,000 function evaluations. For the GA the first point on the plot is added at the end of the first generation, when 64 flow solutions have already been computed (population size is 64). The same is true for the HM except that at the end of the first generation 775 flow solves have been performed. This explains the gaps before the first points for GA and HM on the convergence plots. For GB-MS the first Sobol point converges to the global optimum; the convergence plot does not reflect the total number of samples considered.

The results indicate that the design space is unimodal, and the local optimum found using the RAE2822 as a starting point is the global optimum. We have performed similar studies varying the following parameters: 1) different numbers of design variables, 2) different Mach numbers ($Ma = 0.50$ and $Ma = 0.85$), and 3) a different objective function ($C_d/C_l$).

Only one local minimum was found in each case. Therefore, our conclusion is that this is a unimodal problem, and the most effective algorithm for this type of optimization problem is GB. However, using GB-MS is still useful to gain confidence that no additional local optima exist.

### D. Transonic Wing-Section Optimization

The 3-D ASO problems are solved using the inviscid Newton-Krylov-Schur solver (see Sec. II). In all wing optimization cases the mesh is a 12-block structured H-H topology grid with a total of 1,158,300 nodes, as shown in Fig. 9. The length units are nondimensionalized with the initial chord length being equal to 2/3.
The wing is parameterized using a 2-patch B-spline surface. The baseline initial guess has NACA0012 cross-sections and a rectangular planform with a semispan of two. Projected area of the half-wing is used as a reference and is constrained to 4/3 using a nonlinear constraint.

We perform an optimization of the wing sections with NACA 0012 cross-sections as the baseline. The objective is to minimize $CD_{Ma/0.80}$. The lift coefficient is constrained to $CL_{Ma/0.80} = 0.2625$. The volume is constrained to be at least $6.57 \times 10^{-2}$, which is the volume of the original wing.

Each patch contains five control points in the streamwise direction and six in the spanwise direction. The control point at the root trailing edge is fixed. The control points at the root are not allowed to move in the spanwise direction. Accounting for the duplicate control points on the stitches there are 125 geometric design variables (control-point coordinates). The angle of attack is also a design variable. For this problem we are only optimizing cross sections thus allowing the interior control points to move only in the vertical direction. This is accomplished using linear constraints as described in Sec. III.

The solution of the baseline geometry is plotted in Fig. 10a where the Mach number contours are shown. There is a relatively strong shock on the upper surface and $CD_{Ma/0.80} = 3.158 \times 10^{-2}$.

The GB optimization took 57 major iterations (60 function evaluations) to converge to within $10^{-8}$ optimality tolerance, which demonstrates that a local optimum has been reached. $CD_{Ma/0.80}$ at the optimal point is $3.582 \times 10^{-3}$. In Fig. 10b we can see that the final geometry is shock free. Figure 11 shows that the spanwise lift distribution is nearly elliptical, except near the tip, where the wing deformation is not allowed by the linear constraints, and side edge separation can be present. Therefore, the optimizer has successfully eliminated the shock and minimized induced drag within the constraints. The convergence plot for the GB optimizer is shown in Fig. 12a.

For GB-MS method 128 initial geometries, determined by the Sobol sequence, were optimized. No additional local optima were found for this problem and all optimizers converge to the same point. For the HM optimization the population size is 64, and the algorithm was run for 10 generations. The local optimum was reached on the second generation. Subsequent generations failed to find any additional local optima.

The convergence plots for all optimizers are provided in Fig. 12b. Because only one local optimum was found in this problem, we conclude that the design space is unimodal, and the GB method is the most effective for this type of ASO problem.

### E. Subsonic Wing Optimization

In this problem we consider subsonic flow and allow greater geometric flexibility than in the previous problem. This is achieved by changing the linear constraints while the number and location of the design variables remain identical to the transonic wing-section
optimization. The objective is to minimize the drag coefficient $C_D$. The Mach number is 0.50. The lift coefficient $C_L$ is constrained to 0.2625. The projected area is constrained to $4/3$. The angle of attack is $6\,\text{deg}$. The Mach number is 0.50. The lift coefficient $C_L$ is constrained to $4/3$. The volume is constrained to $6.57 \times 10^{-2}$, which is the volume of the baseline geometry. The control point at the trailing edge can have a maximum spanwise extent of 2.4, maximum sweep back to 1.00 (from the original value of 0.33), and vertical bounds are $-0.3$ and 0.3. Each section is allowed to twist and change its shape. Both leading and trailing edges can be curved. The first 32 initial geometries generated by the Sobol sequence are illustrated in Fig. 13.

Because the flow is subsonic and inviscid, only induced drag is present. The baseline initial guess has $C_D = 4.094 \times 10^{-3}$. Gradient-based optimization took 257 major iterations. After the optimization process $C_D$ was reduced to $1.744 \times 10^{-3}$. The convergence plot is provided in Fig. 14a, and the optimized wing geometry is shown in Fig. 15. It is worth noting that the geometries shown in Fig. 15 are not representative of practical wing shapes. The primary objective here is to investigate the existence of multiple local optima. In the absence of structural constraints one must define further geometric constraints to produce more practical wing shapes (e.g., make trailing edge straight, etc.). Our experience suggests that defining such constraints further reduces multimodality [15,16].

For GB-MS method 192 initial guesses, determined by the Sobol sequence, were optimized. The optimization process found seven distinct local optima all of which converged to the optimality tolerance of $10^{-6}$. These optimal geometries are shown in Fig. 15. $C_D$ values for all local optima found are compared in Table 2. The results show that the baseline geometry does not lead to the global optimum, and that a further 0.54% drag improvement can be obtained by considering alternative initial geometries. It is also important to note the significant difference in performance at the local optima. The local optima differ by more than 5% in the objective value.

For the hybrid optimizer the population size is 64. Each population member has a SNOPT major iteration limit of 50. The HM optimizer was run for 10 generations. As one can see in Fig. 14b HM is able to find the same best optimal point as the GB-MS, but it takes more function evaluations to reach that point. The GA was run for 100 generations. As shown in Fig. 14b, the GA converges much more slowly than the other algorithms.

Because seven local optima were found in this problem, we conclude that the design space is somewhat multimodal. Increasing the Mach number to 0.80 results in three local optima, which falls under the same multimodality classification. In both subsonic and transonic cases the GB-MS optimizer is the most efficient algorithm.

F. Comments on Grid Refinement

Because induced drag is particularly sensitive to grid size, it is important to understand how the results are affected as the grid is refined. We address two questions:
1) Do the optimized geometries change with the refined mesh?
2) Do the $C_D$ values change with the refined mesh?

To answer the first question we increased the number of nodes in the original mesh by a factor of two in all directions. The resulting grid has 12 blocks and 8,955,180 nodes. The optimization process was started from the local optima found using the original grid size of 1,158,300. The optimizer is able to converge these geometries to an optimality tolerance of $10^{-6}$, so local minima are again found. The optimized geometries on the fine grid are very similar to the geometries on the coarse grids. Therefore, it appears that the optimal shapes are grid independent.

To answer the second question we increased the number of nodes in the original mesh by a factor of four in all directions. Each block was split into 64 blocks. The resulting grid has 70,416,972 nodes with 768 blocks. Flow solutions were computed for three of the optimized geometries on this grid. The values of $C_L$, $C_D$, and the span efficiency
factor $e$ are compared in Table 3. The values of $C_D$ change significantly as the grid is refined. Because the values of $C_L$ change slightly, the relative performance of these wing geometries is best assessed in terms of the span efficiency factor $e$. The relative performance at the local optima remains consistent (i.e., optimized geometries are superior to the baseline geometry, local optimum 2 is better than local optimum 7, etc.).

G. Blended-Wing-Body-Configuration Optimization

A Blended-wing-body (BWB) is an unconventional configuration that has the potential to reduce fuel burn by as much as 30% compared to a conventional wing-tube configuration [30]. In this problem we apply the optimization algorithms to a BWB configuration.

<table>
<thead>
<tr>
<th>Geometry</th>
<th>$C_D$</th>
<th>Relative difference from local optimum 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>$4.094 \times 10^{-3}$</td>
<td>157.73%</td>
</tr>
<tr>
<td>Local optimum 1</td>
<td>$1.588 \times 10^{-3}$</td>
<td>0.00%</td>
</tr>
<tr>
<td>Local optimum 2</td>
<td>$1.580 \times 10^{-3}$</td>
<td>−0.54%</td>
</tr>
<tr>
<td>Local optimum 3</td>
<td>$1.634 \times 10^{-3}$</td>
<td>2.84%</td>
</tr>
<tr>
<td>Local optimum 4</td>
<td>$1.621 \times 10^{-3}$</td>
<td>2.04%</td>
</tr>
<tr>
<td>Local optimum 5</td>
<td>$1.652 \times 10^{-3}$</td>
<td>4.00%</td>
</tr>
<tr>
<td>Local optimum 6</td>
<td>$1.644 \times 10^{-3}$</td>
<td>3.41%</td>
</tr>
<tr>
<td>Local optimum 7</td>
<td>$1.669 \times 10^{-3}$</td>
<td>5.06%</td>
</tr>
</tbody>
</table>

The geometry is parameterized with four B-spline patches. Each patch has seven control points in the streamwise direction and six in the spanwise direction. The total number of design variables is 368. A system of linear constraints was created to control the geometric

---

Table 3 Grid-refinement results

<table>
<thead>
<tr>
<th>Quantity</th>
<th>$C_L$</th>
<th>$C_D$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesh Size</td>
<td>1,158,300</td>
<td>70,416,972</td>
<td>1,158,300</td>
</tr>
<tr>
<td>Baseline Geometry</td>
<td>0.2625</td>
<td>0.2659</td>
<td>$4.094 \times 10^{-3}$</td>
</tr>
<tr>
<td>Local optimum 1</td>
<td>0.2625</td>
<td>0.2639</td>
<td>$1.588 \times 10^{-3}$</td>
</tr>
<tr>
<td>Local optimum 2</td>
<td>0.2625</td>
<td>0.2623</td>
<td>$1.580 \times 10^{-3}$</td>
</tr>
<tr>
<td>Local optimum 7</td>
<td>0.2625</td>
<td>0.2611</td>
<td>$1.669 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
deformation. The mesh, shown in Fig. 16, is an H-H topology grid with 627,000 nodes. The drag-coefficient value for the baseline geometry is $C_D = 1.849 \times 10^{-2}$. The top planform view with the Mach-number contours is displayed in Fig. 17a showing a strong shock on the upper surface.

The objective is to minimize $C_D$ with $C_L$ constrained to 0.3522. The projected area is constrained to 0.24133, and the Mach number is 0.80. We allow variation in the sweep, dihedral, and linear twist of the wing as well as section changes on both wing and body. Leading and trailing edges are straight. The span is constrained to remain at its initial value.

Even with these constraints the design space allows for very large variations in geometry. Not all of these geometries can be handled by the flow solver especially at transonic speed. For the GB optimizer a geometry with $C_D$ value of $1.230 \times 10^{-2}$ is produced. This geometry satisfies all of the constraints, but the optimality criterion is not reduced to within the set tolerance of $10^{-6}$. Table 4 shows the performance of the baseline geometry and this optimized geometry, which is identified as local optimum 1; its planform is displayed in Fig. 18. The drag coefficient of the baseline geometry is over 50% higher than that of local optimum 1.

Although 224 initial geometries were considered for GB-MS optimizer, only 33 initial geometries converged on the first iteration.

Table 4  Results summary for geometries in BWB problem

<table>
<thead>
<tr>
<th>Geometry</th>
<th>$C_D$</th>
<th>% Difference from Local optimum 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>$1.849 \times 10^{-2}$</td>
<td>50.3%</td>
</tr>
<tr>
<td>Local optimum 1</td>
<td>$1.230 \times 10^{-2}$</td>
<td>0.00%</td>
</tr>
<tr>
<td>Local optimum 2</td>
<td>$1.289 \times 10^{-2}$</td>
<td>4.81%</td>
</tr>
<tr>
<td>Local optimum 3</td>
<td>$1.244 \times 10^{-2}$</td>
<td>1.09%</td>
</tr>
<tr>
<td>Local optimum 4</td>
<td>$1.241 \times 10^{-2}$</td>
<td>0.87%</td>
</tr>
<tr>
<td>Local optimum 5</td>
<td>$1.249 \times 10^{-2}$</td>
<td>1.54%</td>
</tr>
<tr>
<td>Local optimum 6</td>
<td>$1.242 \times 10^{-2}$</td>
<td>0.85%</td>
</tr>
<tr>
<td>Local optimum 7</td>
<td>$1.244 \times 10^{-2}$</td>
<td>1.12%</td>
</tr>
<tr>
<td>Local optimum 8</td>
<td>$1.258 \times 10^{-2}$</td>
<td>2.25%</td>
</tr>
</tbody>
</table>
allowing the optimization process to continue. From these 33 initial geometries eight distinct local optima are identified that satisfy both optimality and feasibility tolerances of $10^{-6}$. The objective values for these local optima are summarized in Table 4. One can see that the objective value can vary by almost 5% between the different local optima. It is also worth noting that the GB optimizer was able to find the best geometry in terms of the value of the objective function. We can regard it as a random occurrence that the baseline geometry leads the optimizer to the shape with the lowest drag coefficient. As we have seen in the previous problem this is not always true. We can also expect that the GB-MS optimizer will eventually find the same local optimum if more initial samples are considered.

The planform shapes of the local optima are shown in Fig. 18. Clearly, all shapes are distinct, demonstrating that the shape variation among the local optima can be significant. The convergence plot for local optimum 4 is shown in Fig. 19a, which shows that a local optimum has been reached. Figure 17 compares the Mach-number-contour plots on the upper surfaces of the baseline geometry and geometry corresponding to local optimum 4. The optimizer is able to eliminate the shock on the upper surface.

For completeness we show convergence plots for all four optimization methods in Fig. 19b. As in the previous problems one can see that the GA optimizer converges more slowly than the other optimizers. Due to time constraints the HM optimizer was allowed to run for five generations only, and it can be observed that its convergence rate is comparable to the GB-MS method. As with the GB-MS method, it is expected that the HM will eventually converge to the geometry found by the GB optimizer.

This study is only preliminary and further investigation is required to assess the performance of the BWB configuration. We acknowledge that the coarse mesh used for this problem can introduce significant numerical error. The main purpose here is to underscore the importance of considering a global optimization method. It appears that this problem is at least somewhat multimodal and possibly moderately multimodal. The results indicate that GB-MS is an efficient method for finding the global optimum of high-fidelity ASO problems that contain a substantial number of local optima, but as the number of local optima increases, the HM eventually becomes the preferred method.

V. Conclusions

The results show the shortcomings of using either pure gradient-based or gradient-free optimization algorithms for high-fidelity aerodynamic shape optimization. A purely gradient-based optimizer may converge to a local optimum with an objective value significantly worse than that of the global optimum. On the other hand, the computational expense of using a gradient-free optimization algorithm is unacceptable for many problems especially if the number of design variables is large.

Two new optimization algorithms are presented to address these shortcomings. The algorithms are efficient for global optimization of smooth differentiable objective functions with multiple local optima. The linear constraints allow the size of the problem to be reduced and ensure only feasible geometries are considered. The sampling method uses a Sobol sequence to cover the design space in a uniform and unbiased manner.

The optimization results show that multimodality is not a predominant feature of the design spaces for many aerodynamic and turbulent flows and conventional configurations. However, multiple local optima do exist, and a global optimization method is essential to find the most efficient aerodynamic configuration.

Airfoil optimization and wing section optimization problems are unimodal, and a gradient-based method is preferred for optimization, although the use of the gradient-based multistart procedure is informative to ensure that no other local optima exist. Wing-shape optimization problems that involve considerable geometric deformations are somewhat multimodal, and the gradient-based multistart method is most efficient. The blended-wing-body optimization problem also appears to be somewhat to moderately multimodal, although further investigation is required to establish the degree of multimodality of this case. The number of local optima tends to increase with increasing geometric flexibility. As the number of local optima increases the hybrid optimization algorithm method eventually becomes more efficient than the gradient-based multistart method. The results presented suggest that a local optimization algorithm such as a gradient-based algorithm will often be sufficient for high-fidelity aerodynamic shape optimization in detailed design, where geometric flexibility is limited. In contrast, a global optimization algorithm, such as those presented in this paper, is needed for exploratory high-fidelity aerodynamic shape optimization, where large shape changes are permitted.

Future work will include testing the multimodality of aerodynamic shape optimization problems in viscous and turbulent flows and further studies of unconventional aerodynamic configurations. Adding structural requirements for high-fidelity aerostructural optimization will also be considered in terms of the impact on multimodality.

Acknowledgments

The authors would like to thank T. Pulliam for help with the genetic algorithm and F. Kuo for providing directional numbers for Sobol sequences.

Computations were performed on the General Purpose Cluster (GPC) supercomputer at the SciNet High Performance Computing (HPC) Consortium. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario, Ontario Research FundResearch Excellence, and the University of Toronto.

The funding of the first author by Mathematics of Information Technology and Complex Systems (MITACS) and Bombardier Aerospace and the second author by the Natural Sciences Research
Council of Canada and the Canada Research Chairs program is gratefully acknowledged.

References


R. Kapania
Associate Editor