

# Improvements to a Newton-Krylov Adjoint Algorithm for Aerodynamic Optimization

David W. Zingg,\* Timothy M. Leung,† Laslo Diosady‡ Anh H. Truong,† and Samy Elias†

*Institute for Aerospace Studies, University of Toronto  
4925 Dufferin St., Toronto, Ontario M3H 5T6, Canada*

Marian Nemec §

*NASA Ames Research Center, Moffett Field, CA 94035*

We present and examine a number of improvements to a gradient-based algorithm for aerodynamic optimization. A Newton-Krylov algorithm is used to solve the compressible Navier-Stokes equations, the gradient is computed using the discrete-adjoint method with a preconditioned Krylov solver, and the optimum is found through a quasi-Newton algorithm with a rank-two update formula. Constraints are imposed by penalizing the objective function. Improvements are made in three areas: 1) thickness constraints are generalized to permit the location of maximum thickness to be determined by the optimizer or alternatively to constrain the cross-sectional area; 2) new scalings of design variables and initial estimates of the inverse Hessian matrix in the quasi-Newton method are investigated; 3) the algebraic grid perturbation algorithm is replaced by an algorithm based on a spring analogy. In each case, the effect of the improvements on the performance of the algorithm is presented.

## I. Introduction

Numerical aerodynamic shape optimization can be of great use in the design process. It relieves the designer of the tedious task of searching the design space and thereby enables added focus on the specification of the design problem, i.e. the definition of the design space, the objectives, and the constraints. The adjoint method, whether applied in a discrete or a continuous manner, has proven to be an efficient approach to aerodynamic optimization when the number of design variables is large.<sup>1,2</sup>

Nemec and Zingg<sup>3</sup> presented an efficient Newton-Krylov algorithm for aerodynamic shape optimization based on the discrete adjoint method. The algorithm provides the basis for the two-dimensional aerodynamic optimization codes *Optima2D*, for single-element airfoils, and *OptimaMB*, for multi-element configurations. Their approach includes all terms needed to calculate accurate gradients, such as grid sensitivities and the linearization of the coupled turbulence model. As a result, the gradient is typically reduced by five orders of magnitude, ensuring convergence to at least a local minimum. While this can be more expensive than other approaches, a high degree of convergence is important when assessing design trade-offs. Alternatively, if quick turnaround is a high priority, one can run a fixed number of iterations, since the majority of the overall improvement in the objective function typically occurs in a fraction of the total number of iterations needed to converge fully (although the same cannot be said of the shape changes). The formulation of Nemec and Zingg has been applied to numerous design problems, including multipoint and multi-objective problems<sup>4</sup> and high-lift configurations.<sup>5</sup>

The objective of this paper is to describe and demonstrate three improvements to *Optima2D*. The first provides more flexibility in the specification of geometric constraints. The second modifies the scaling of

---

\*Professor, Senior Canada Research Chair in Computational Aerodynamics, 2004 Guggenheim Fellow, Senior AIAA Member.

†Graduate Student

‡Undergraduate Student

§NRC Research Associate, presently Research Scientist, ELORET Corp., Applications Branch, Advanced Supercomputing Division, AIAA Member.

the design variables and the initial estimate of the Hessian in the quasi-Newton method in order to provide faster and more reliable convergence. The final improvement is a more robust procedure for perturbing the grid in response to shape changes. Examples are provided showing the benefits of each improvement.

## II. Algorithm Description

The present algorithm parameterizes the geometry using B-splines. The B-spline control points and the angle of attack are the design variables. The compressible Navier-Stokes equations are solved with a Newton-Krylov method in which the linear system arising at each Newton iteration is solved using the generalized minimal residual method (GMRES) preconditioned with an incomplete lower-upper (ILU) factorization with limited fill. The Spalart-Allmaras turbulence model is used to compute the eddy viscosity. The gradient is calculated using the discrete adjoint method; solution of the adjoint equation is accomplished through the same preconditioned Krylov method. Geometric constraints are added to the objective function as penalty terms. A new set of design variables is computed using a quasi-Newton optimizer in which an estimate of the inverse Hessian based on the BFGS (Broyden-Fanno-Goldfarb-Shannon) rank-two update formula is used to compute a search direction.<sup>6</sup> If the initial step does not produce sufficient progress toward the minimum, the step size is determined using a line search, which terminates when the strong Wolfe conditions are satisfied.<sup>6</sup> Each time a new shape is calculated, the initial grid is perturbed using a simple algebraic technique. For a complete description of the algorithm, see Nemeč.<sup>7</sup>

## III. Algorithm Improvements

### A. Geometric Constraints

In aerodynamic shape optimization, geometric constraints are typically introduced to represent structural and manufacturing requirements as well as practical considerations, such as fuel storage. In *Optima2D*, the designer can specify a minimum thickness at any location along the airfoil chord. Typically several such thickness constraints are introduced both for the purposes listed above and to prevent crossover during the optimization iterations. The latter are generally inactive once a converged shape is obtained. While this approach provides the designer with a great deal of control, it can sometimes be preferable to allow more flexibility to the optimizer. For example, rather than specifying a thickness to chord ratio of say 12% at 35% chord, it may be sufficient to require simply that the thickness to chord ratio be 12% at some unspecified chord location. Even more flexibility is provided if the geometric constraint is expressed in terms of a minimum cross-sectional area, rather than a minimum thickness.

### B. Scaling and the Initial Inverse Hessian Estimate

The quasi-Newton algorithm determines a search direction based on the gradient and an approximation to the inverse of the Hessian matrix. The first search direction is equivalent to the steepest-descent direction. After the first step, an initial estimate of the inverse Hessian is required. The BFGS secant update is then used to generate a sequence of matrices that approximate the inverse of the exact Hessian with increasing accuracy.

The convergence of the quasi-Newton method can be affected by the scaling of the design variables.<sup>8,9</sup> Although the scaling does not affect the BFGS updates themselves, it influences the first step and therefore the initial estimate of the inverse Hessian, and this influence persists throughout the optimization. A poor initial inverse Hessian leads to poor inverse Hessian estimates in subsequent iterations, causing slow convergence and possibly line-search stall. Nemeč and Zingg scaled each design variable by its initial value with a correction if the initial value is zero.<sup>3</sup> Hence the scaled design variables  $x'$  are related to the unscaled design variables  $x$  by the following equation:

$$x' = L^{-1}x \tag{1}$$

where  $L$  is a diagonal matrix containing the initial values of the unscaled design variables. The ideal scaling is related to the square root of the diagonal entries of the Hessian matrix and is not directly related to the initial values of the design variables. However, the Hessian is of course unknown.

The initial estimate of the scaled inverse Hessian is currently given by:<sup>6</sup>

$$H'_0 = \frac{v'^T d'}{v'^T v'} I \quad (2)$$

with

$$\begin{aligned} v' &= g'_1 - g'_0 \\ d' &= x'_1 - x'_0 \end{aligned}$$

where the prime denotes scaled variables, and  $g$  is the gradient vector. An inappropriate scaling can lead to a poor initial estimate of the inverse Hessian from which the BFGS updates cannot recover. Consequently, it can be advantageous to use the unscaled estimate of the inverse Hessian.

Whether the scaled or unscaled initial estimate is used, these estimates are simply scalar multiples of the identity matrix. Alternatively, we consider an initial estimate of the inverse Hessian which is a diagonal matrix with entries given by:

$$H_{ii} = \frac{d_i}{v_i} \quad (3)$$

A fix is needed if any of the entries has a value below a certain threshold.

### C. Grid Perturbation

The algebraic grid perturbation technique has proven to be effective for a range of optimization problems. For problems such as finding the optimum position of a flap, however, it can lead to grids of insufficient quality, and periodic regridding can be required. Regridding often involves human intervention, especially in three dimensions, and is thus undesirable. Poor quality grids can also result from the simple algebraic perturbation technique when the shape changes are substantial. For example, if the angle of the trailing-edge bisector changes too much, grid distortion can be excessive.

In order to avoid regridding, we consider an alternative grid perturbation technique based on the lineal spring analogy.<sup>10,11</sup> The purpose of a grid perturbation technique is to propagate the deformation of the boundary, or more specifically, the surface grid, into the interior of the grid while maintaining grid quality. In the lineal spring method, the edges of the elements act as springs whose stiffness is proportional to the edge length, i.e.

$$k_{ij} = 1 / \left[ (x_j - x_i)^2 + (y_j - y_i)^2 \right]^{\frac{1}{2}} \quad (4)$$

The original grid is in equilibrium. When the surface nodes move in response to a shape change, forces are generated due to the displacement. The force at node  $i$  is

$$F_i = \sum_{j=1}^{N_i} k_{ij} (\delta_j - \delta_i) \quad (5)$$

where  $N_i$  is the number of neighbouring nodes of node  $i$ , and  $\delta_i$  is the displacement of node  $i$  from its position in the original grid. The interior nodes must thus move such that equilibrium is restored, i.e.  $F_i = 0$ . This leads to a linear system of equations, which is solved using the conjugate gradient method. Since the grid cells near the airfoil surface tend to be small, they have high spring stiffness, and thus their deformation is minimal.

A key aspect of this procedure in the present context is that the grid perturbation is always carried out relative to the original grid around the initial shape, not relative to the grid generated from the previous optimization iteration. Therefore, there is a unique grid and hence a unique solution for a given set of design variables. This is an important property for an optimization algorithm.

## IV. Results

### A. Geometric Constraints

The geometric constraints are studied in the context of lift-to-drag ratio maximization at a Mach number of 0.75 and a Reynolds number of 2.88 million. Fully turbulent flow is assumed, i.e. transition is assumed to



Figure 1. B-spline control points; the shaded control points are used as design variables.

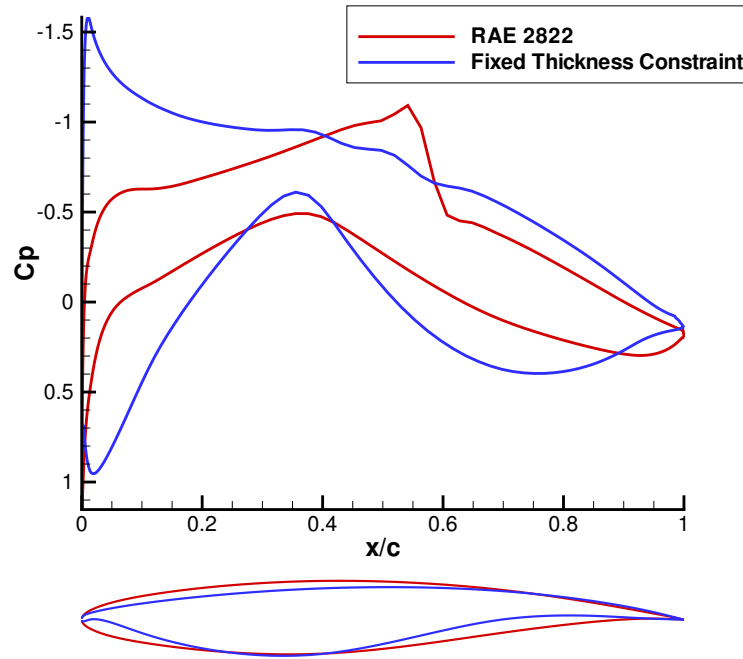


Figure 2. Surface pressure coefficient distribution and airfoil shapes: RAE 2822 airfoil and optimum airfoil with fixed thickness constraint at  $x/c = 0.35$ .

occur at the leading edge. The initial geometry is the RAE 2822 airfoil. Seventeen B-spline control points are used to parameterize the geometry. Three control points are frozen at the leading edge and four at the trailing edge. Hence there are eleven design variables including the angle of attack. Figure 1 shows the B-spline control points. The grid has a “C” topology with 289 nodes in the streamwise direction and 65 in the normal direction.

For the first example, two thickness constraints are applied. A thickness to chord ratio of at least 11% is imposed at  $x/c = 0.35$ , and a 0.2% thickness to chord ratio is imposed at  $x/c = 0.99$ . The latter constraint is introduced to prevent crossover near the trailing edge. The pressure distributions and airfoil shapes are shown in Figure 2 for the initial (RAE 2822) and final airfoils. The shock wave has been eliminated on the upper surface. The resulting airfoil has lift and drag coefficients of 0.8195 and 0.0147, respectively, at an angle of attack of 2.56 degrees, for a lift-to-drag ratio of 55.31.

The same optimization problem is repeated with a minimum thickness requirement of 11%, but no particular value of  $x/c$  is specified. The resulting airfoil (labelled “floating thickness constraint”) is compared with the previous airfoil in Figure 3. The location of maximum thickness has moved slightly aft. The resulting lift and drag coefficients are 0.8286 and 0.0148, respectively, at an angle of attack of 2.72 degrees, giving a lift to drag ratio of 55.84. The optimizer has exploited the additional freedom to improve the lift to drag ratio by roughly 1%. The benefit is relatively small because the specified location of the maximum thickness is not far from the optimum location of maximum thickness. However, in other problems, for example in the

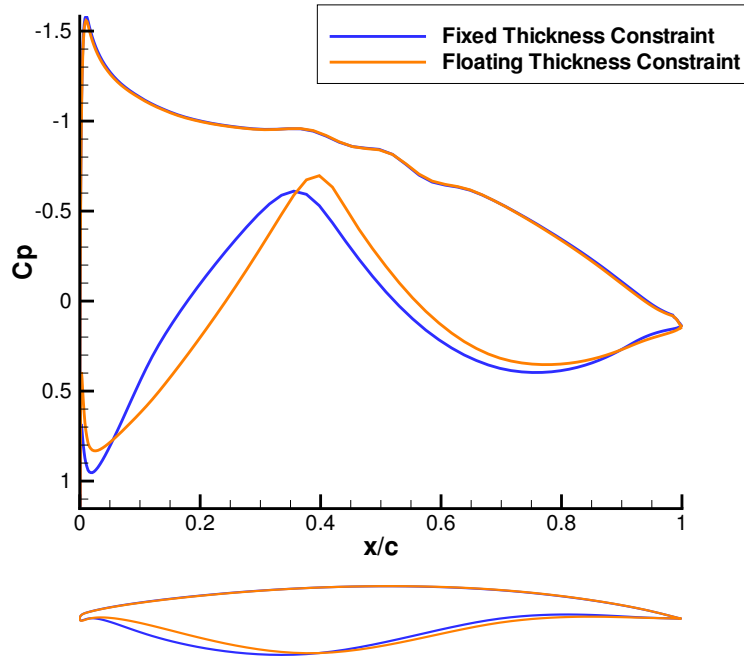


Figure 3. Surface pressure coefficient distribution and airfoil shapes: optimum airfoil with fixed thickness constraint at  $x/c = 0.35$  and optimum airfoil with floating thickness constraint.

design of natural-laminar-flow airfoils, the location of maximum thickness may be difficult to guess. Unless there is a specific reason to constrain the location of maximum thickness, such as the presence of a spar, it is preferable not to do so.

Finally, the same optimization is repeated without any thickness constraints. The area is required to be equal to the cross-sectional area of the airfoil generated using the floating thickness constraint. This leads to lift and drag coefficients of 0.8368 and 0.0141, respectively, at an angle of attack of 2.55 degrees, resulting in a lift to drag ratio of 59.22. This is roughly 6% higher than the lift to drag ratio of the airfoil designed using the floating thickness constraint. Furthermore, the resulting shape is smoother with less pinching near the leading edge and a more uniform load distribution, as shown in Figure 4.

## B. Scaling and the Initial Inverse Hessian Estimate

Two optimization problems are used to examine the effect of the design-variable scaling and the initial inverse Hessian estimate. The first is a lift-constrained drag minimization with  $C_l = 0.733$ . The second is a maximization of the lift-to-drag ratio. The following parameters are common to both optimization problems:

- $M = 0.74$ ,  $Re = 2.7$  million
- 30 B-spline control points with two frozen at the leading edge and two frozen at the trailing edge; 27 design variables including the angle of attack
- thickness constraints:  $t/c \geq 0.1204$  at  $x/c = 0.35$ ,  $t/c \geq 0.005$  at  $x/c = 0.96$ ,  $t/c \geq 0.0012$  at  $x/c = 0.99$
- $257 \times 57$  grid

Figure 5 shows the converged airfoil and surface pressure coefficient distribution for the lift-constrained drag minimization, which has  $C_l = 0.7237$ ,  $C_d = 0.01432$ . The corresponding values for the lift-to-drag ratio maximization are  $C_l = 0.9354$ ,  $C_d = 0.01731$ , for a lift-to-drag ratio of 54.

The following four design-variable scalings have been studied:

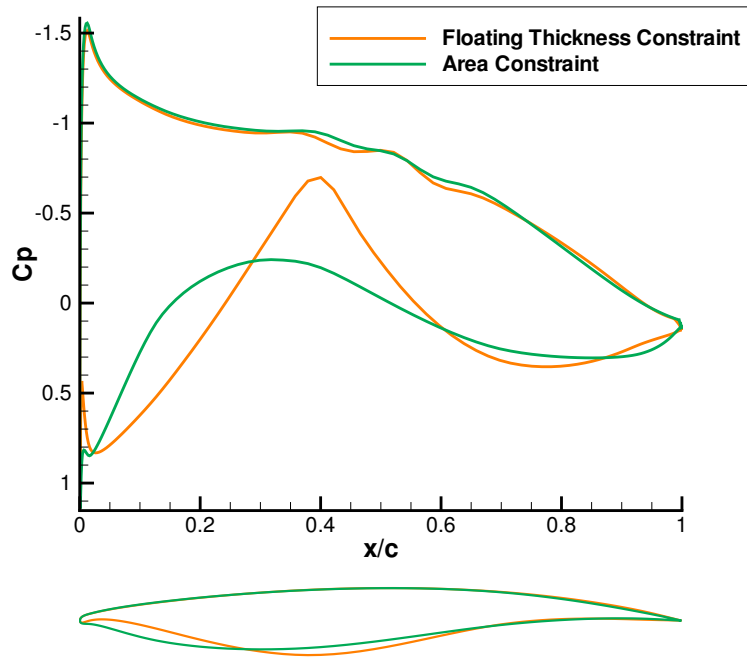


Figure 4. Surface pressure coefficient distribution and airfoil shapes: optimum airfoil with floating thickness constraint and optimum airfoil with area constraint.

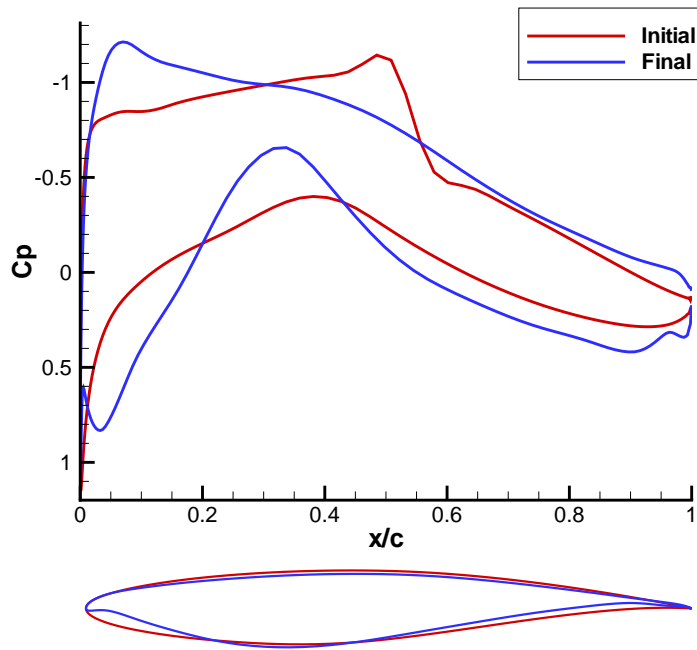


Figure 5. Surface pressure coefficient distribution and airfoil shape for lift-constrained drag minimization.

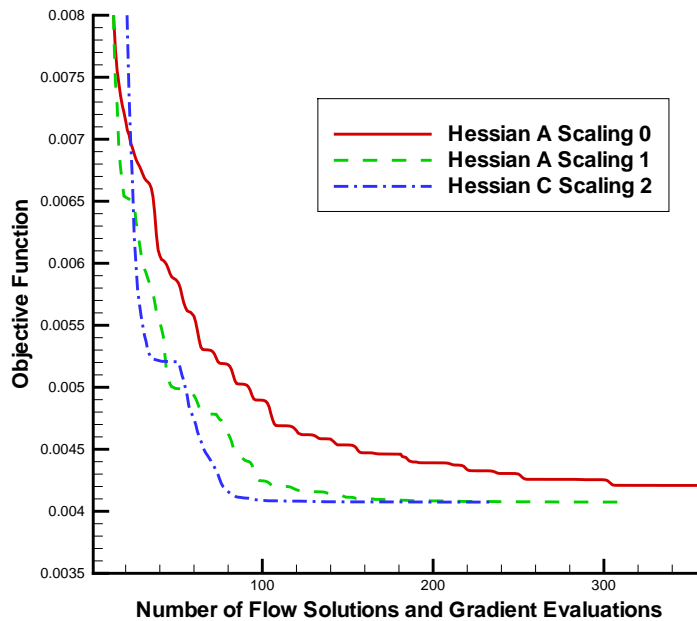


Figure 6. Convergence histories for the lift-constrained drag minimization problem.

- Scaling 0: each design variable is scaled by its initial value (with a fix if the initial value is zero)
- Scaling 1: the angle of attack design variable is scaled by its initial value; the B-spline control point design variables are scaled by an average value
- Scaling 2: each design variable is scaled by the square root of its initial value (with a fix if the initial value is zero)
- Scaling 3: no scaling is applied

Scaling 2 is based on the observation that Scaling 0 sometimes provides too much scaling, while Scaling 3 often provides too little.

In addition, the following three initial inverse Hessian estimates have been examined:

- Hessian A: the original approach given by Eq. 2
- Hessian B: similar to Hessian A, but calculated using *unscaled* variables
- Hessian C: the diagonal matrix given by Eq. 3

All possible combinations of scalings and initial inverse Hessian estimates have been applied to numerous optimization problems. The most promising combinations are Hessian A with Scaling 1 and Hessian C with Scaling 2. Convergence histories for these two combinations are compared with the baseline scheme, Hessian A with Scaling 0, in Figures 6 and 7. The lift-constrained drag minimization problem is shown in Figure 6; the lift-to-drag ratio maximization is shown in Figure 7.

The results for the two problems are quite different. For the lift-constrained drag minimization, the Hessian C-Scaling 2 combination converges fastest, requiring 96 iterations to converge to within 0.5% of the final converged objective function value. The Hessian A-Scaling 1 combination requires 174 iterations to achieve the same degree of convergence.<sup>a</sup> To converge to within 0.1% of the final objective function value,

<sup>a</sup>Note that an iteration involves one flow solution and one gradient evaluation.

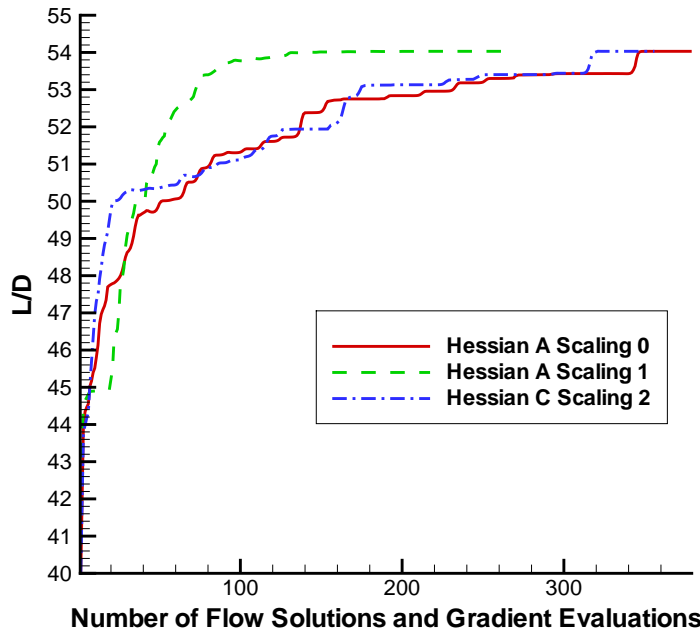


Figure 7. Convergence histories for the lift-to-drag ratio maximization problem.

the two schemes require 132 and 293 iterations, respectively. The baseline scheme stalls without converging and is much slower than the other two combinations to achieve partial convergence. For the lift-to-drag ratio maximization, the combination of Hessien A and Scaling 1 converges fastest, requiring 97 and 137 iterations to achieve 0.5% and 0.1% convergence levels, respectively. The Hessien C-Scaling 2 combination requires 317 and 318 iterations, respectively, while the original Hessien A-Scaling 0 combination requires 345 and 346.

The savings associated with the new schemes relative to the baseline scheme are substantial. For the lift-constrained drag minimization, the Hessien C-Scaling 2 combination converges almost twice as fast as the Hessien A-Scaling 1 combination, and the Hessien A-Scaling 0 combination does not converge. For the lift-to-drag ratio maximization, the combination of Hessien A with Scaling 1 is two to three times faster than both of the other options. Similar results are obtained for both objective functions with different numbers of design variables. Clearly, there is a significant benefit to choosing the best scaling and initial inverse Hessien estimate for a given optimization problem. The difficulty is that at this point we do not understand why these combinations behave differently for the two objective functions. Therefore, pending further study, we recommend the Hessien C-Scaling 2 combination for lift-constrained drag minimization problems and the Hessien A-Scaling 1 combination for lift-to-drag ratio maximization.

### C. Grid Perturbation

The first example is an optimization of the flap position in a two-element geometry which was previously studied by Nemeč.<sup>7</sup> The objective is to find the location of the flap that maximizes the lift coefficient without increasing the drag coefficient. The angle of attack and the flap angle are held fixed. Hence there are two design variables, the coordinates of the flap. The initial grid in the vicinity of the main airfoil trailing edge is shown in Figure 8. Figure 9 shows the new flap position and grid obtained using the algebraic grid perturbation technique. The grid is badly skewed in the region between the grid line emanating from the main airfoil trailing edge and the upper surface of the flap. The grid resulting from the spring-analogy technique is displayed in Figure 10. This grid is less skewed and more comparable to the initial grid. We are currently examining the quantitative implications of this apparent improvement in grid quality.

Figure 11 shows another example of a grid generated using the spring-analogy technique with large shape



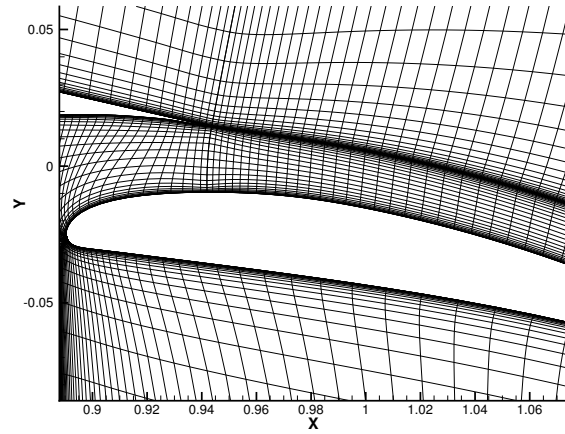


Figure 8. Original grid, showing the region between the trailing edge of the main airfoil and the flap.

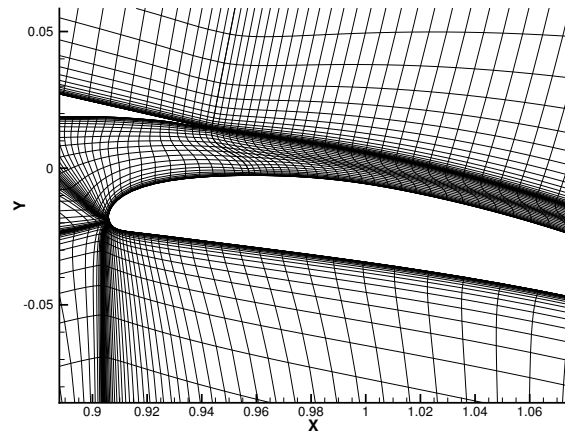


Figure 9. Grid after flap movement using algebraic grid perturbation method.

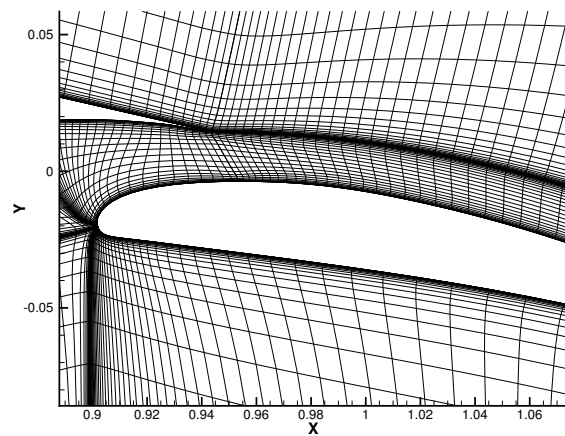
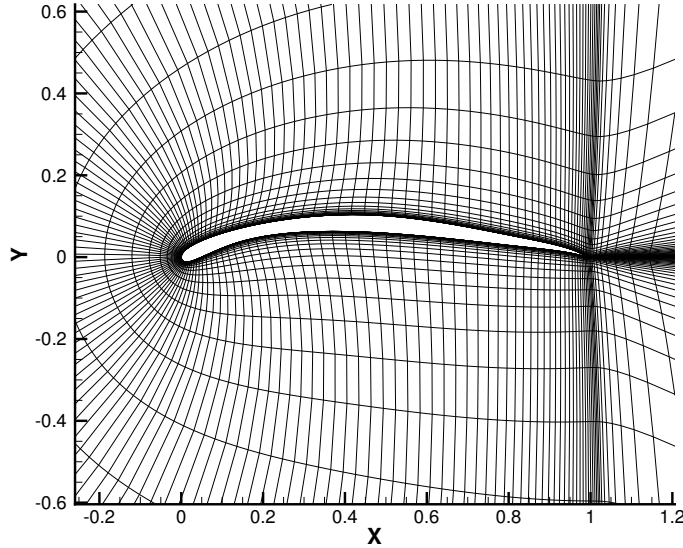


Figure 10. Grid after flap movement using spring-analogy grid perturbation method.



**Figure 11. Spring-analogy grid for the subsonic lift-to-drag ratio maximization problem.**

changes. The initial airfoil is the NACA 0012. The objective is to maximize the lift-to-drag ratio at a Mach number of 0.25 and a Reynolds number of 2.88 million. The following three thickness constraints are imposed: a thickness to chord ratio of at least 4% at  $x/c = 0.05$ , 4% at  $x/c = 0.65$ , and 1.2% at  $x/c = 0.95$ . The final airfoil has a lift-to-drag ratio of 61 under these operating conditions. The grid perturbation technique based on the spring analogy leads to somewhat more reliable convergence without regridding than the algebraic technique for cases such as this with large shape changes.

Figure 12 shows the grid produced using the spring-analogy technique for a lift-to-drag ratio maximization at transonic speed. The initial airfoil is again the NACA 0012, so the shape changes are large, especially on the lower surface near the leading and trailing edges. The Mach number is 0.75, the Reynolds number, 2.88 million. The thickness to chord ratio is constrained to be at least 11% at  $x/c = 0.35$ , 3% at  $x/c = 0.02$ , 2% at  $x/c = 0.85$ , and 0.2% at  $x/c = 0.99$ . In this case, the optimization using the algebraic grid perturbation technique stalled after a small reduction in the gradient, while the spring analogy technique led to a gradient reduction of several orders of magnitude.

These results must be considered preliminary. Although the spring-analogy technique shows promise, we have yet to generate conclusive evidence that it offers a quantifiable advantage over the algebraic technique. This investigation is still underway, and an approach based on a pseudo-elasticity model is also under study.<sup>12</sup>

## V. Conclusions

We have presented improvements to a Newton-Krylov discrete-adjoint algorithm for aerodynamic shape optimization, including more flexible geometric constraints, a new design-variable scaling, a new initial estimate of the inverse Hessian, and an improved grid perturbation technique. Examples show that the area constraint leads to a more efficient, smoother airfoil, while the improvements to the optimizer lead to substantially faster and more reliable convergence.

## Acknowledgments

The funding of the first author by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program is gratefully acknowledged as is the funding of the last author through the National Research Council Research Associateship Award.

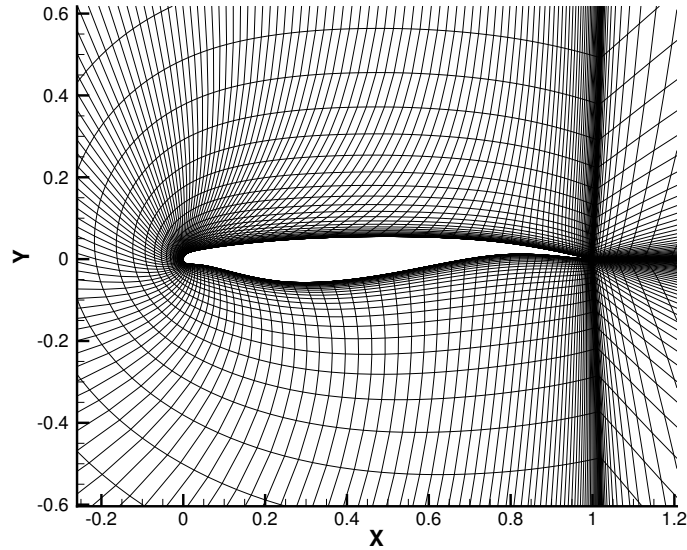


Figure 12. Spring-analogy grid for the transonic lift-to-drag ratio maximization problem.

## References

- <sup>1</sup>Pironneau, O., "On Optimum Design in Fluid Mechanics," *J. Fluid Mech.*, Vol. 64, No. 1, 1974, pp. 97-110.
- <sup>2</sup>Jameson, A., "Aerodynamic Design Via Control Theory," NASA CR-181749, Nov. 1988.
- <sup>3</sup>Nemec, M., and Zingg, D.W., "Newton-Krylov Algorithm for Aerodynamic Design Using the Navier-Stokes Equations," *AIAA J.*, Vol. 40, No. 6, 2002, pp. 1146-1154.
- <sup>4</sup>Nemec, M., Zingg, D.W., and Pulliam, T.H., "Multipoint and Multi-Objective Aerodynamic Shape Optimization," *AIAA J.*, Vol. 42, No. 6, 2004, pp. 1057-1065.
- <sup>5</sup>Nemec, M., and Zingg, D.W., "Optimization of High-Lift Configurations Using a Newton-Krylov Algorithm," AIAA Paper 2003-3957, June 2003.
- <sup>6</sup>Nocedal, J. and Wright, S.J., *Numerical Optimization*, Springer-Verlag, New York, 1999.
- <sup>7</sup>Nemec, M., "Optimal Shape Design of Aerodynamic Configurations: A Newton-Krylov Approach," Ph.D. Thesis, University of Toronto Institute for Aerospace Studies, 2002.
- <sup>8</sup>Gill, P.E., Murray, W., and Wright, M.H., *Practical Optimization*, Academic Press, Toronto, 1981.
- <sup>9</sup>Liu, P.-Y., and Zingg, D.W., "Comparison of Optimization Algorithms Applied to Aerodynamic Design," AIAA Paper 2004-0454, Jan. 2004.
- <sup>10</sup>Batina, J.T., "Unsteady Euler Airfoil Solutions Using Unstructured Dynamic Meshes," *AIAA J.*, Vol. 28, No. 8, 1990, pp. 1381-1388.
- <sup>11</sup>Crumpton, P.I., and Giles, M.B., "Implicit Time-Accurate Solutions on Unstructured Dynamic Grids," *International Journal for Numerical Methods in Fluids*, Vol. 25, 1997, pp. 1285-1300.
- <sup>12</sup>Yang, Z., and Mavriplis, D.J., "Unstructured Dynamic Meshes with Higher-order Time Integration Schemes for the Unsteady Navier-Stokes Equations," AIAA Paper 2005-1222, Jan. 2005.