AIAA JOURNAL
Vol. 38, No. 9, September 2000

Response Surface Techniques for Diffuser Shape Optimization

Jens I. Madsen*
Aalborg University, 9220 Aalborg East, Denmark
and
Wei Shyy† and Raphael T. Haftka‡
University of Florida, Gainesville, Florida 32611-6250

The design of incompressible diffusers for maximum pressure recovery is used to demonstrate the utility of response surface approximations for design optimization of flow devices. Two examples involving two and five design variables are treated, with the diffuser wall shapes described by polynomials and B-splines. In both cases monotonicity conditions drastically reduce the design space. In this irregularly shaped space, a pool of designs is selected by a D-optimality criterion and analyzed by a finite volume computational fluid dynamics (CFD) code. Quadratic polynomial response surfaces are then fitted to the pressure recovery coefficients. To improve the prediction accuracy, uncertain regressor terms and possible outlier design points are excluded based on statistical tests. A standard optimization algorithm is used to find the optimal diffuser design from the response surface approximations. The optimum diffusers exhibit minimal flow separation and yield similar wall shapes for the two parameterizations. A main asset of the response surface optimization approach lies in the smoothing of noisy response functions. Therefore, the issue of numerical noise in CFD results based on the use of two different analysis codes is addressed.

I. Introduction

In step with the rapid development in computer technology, computational fluid dynamics (CFD) has matured to a stage where it provides substantial insight into processes involving engineering fluid flows. However, the primary concern of fluid dynamicists is often not merely to analyze and understand, but moreover to improve performance. This task has traditionally been handled through trial-and-error design procedures, which count on the skills and experience of the designer to suggest redesigns that are likely to yield improvements. While design by trial and error usually yields acceptable solutions, the contemporary use of more rigorous optimization methodologies allows the best design to be identified, with the merit of a design measured explicitly in terms of an objective function. To date, the vast majority of effort in design optimization of fluid machinery has relied on gradient-based search algorithms.[1,2] These methods work iteratively through a sequence of local subproblems, which approximate objective and constraint functions for a subregion of the design space, for example, by linearization using computed first-order design sensitivities. A major challenge of these search optimization approaches is the robust and speedy computation of sensitivity coefficients.[3–5]

Recent years have witnessed an increasing interest in alternative optimization approaches, which avoid the need to compute design sensitivities altogether. The present work explores a promising example thereof, namely, response surface approximations (RSAs) for shape optimization of an internal flow device. Because of a minimum of required software interfacing, which facilitates the integration of several design codes, RSAs are particularly suitable for subsystem approximation within multidisciplinary design optimization.[6] The response surface method has had less impact in single-discipline optimization, such as for fluid design optimization where only a few examples of such work exist (e.g., Refs. 7–9). We focus on the optimum wall contouring of a straight, two-dimensional diffuser with incompressible flow because this is a relatively simple problem of great engineering relevance.

In RSA a series of numerical experiments (here CFD analyses) of appropriately selected designs is conducted, and the sampled response function values are used to construct a global approximation of performance over the design space. The RSA employed here takes the form of quadratic polynomials, which entails fitting by linear regression $[(n + 1)(n + 2)]/2$ coefficients, where n is the number of design variables. Constraint functions that cannot be given explicitly are approximated by a RSA as well. The optimum may then be found at low cost from the response approximations because these are merely algebraic expressions. Depending on the behavior of the objective function near the optimum identified by the algorithm, one can decide whether a refinement of the response surface in this region is worthwhile or not.

The main advantages of optimization via RSA lies in the immediate intelligibility and robustness of the approach. The robustness toward noisy response functions and the smoothness of the approximations follow simply by overfitting the regression model, i.e., using more analyses than regression coefficients. This is a distinct advantage over derivative-based search algorithms, which may encounter substantial difficulties in the presence of spurious local minima. Some further advantages of RSA-based optimization are the suitability of the approach to parallel computation and the useful insight into significance and correlations of individual design parameters given by the regression coefficients of a response surface. Of course, there is cost associated with RSA: its computational demands grow fast with the increasing number of design variables. For example, a quadratic polynomial in 10 variables has 66 coefficients and needs an even higher number of experiments to obtain a sufficiently reliable fit. This strong dependence on problem size has been dubbed the *curse of dimensionality* and is the main reason that derivative-based optimization has remained the preferred approach in single-discipline optimization.

The predictive capability of RSA is greatly influenced by the distribution of sampling points in design space. Therefore, rather than choose designs at random, it is important to apply a sound strategy for selecting these points, a discipline known as the *design of experiments*. Here, design refers to a selection of design points rather than a physical design. We choose to rank different point selections according to the so-called D-optimality criterion, which yields RSAs that are as insensitive to noise as possible.[10] A point

Presented as Paper 97-1801 at the AIAA 28th Fluid Dynamics Conference, Snowmass Village, CO, 29 June–2 July 1997; received 26 May 1999; revision received 18 October 1999; accepted for publication 19 October 1999.

*Research Assistant Professor, Institute of Mechanical Engineering, Pontoppidanstræde 101; jim@ime.auc.dk.

†Professor and Department Chair, Department of Aerospace Engineering, Mechanics and Engineering Science; wei-shyy@ufl.edu. Associate Fellow AIAA.

‡Distinguished Professor, Department of Aerospace Engineering, Mechanics and Engineering Science; haftka@ufl.edu. Fellow AIAA.

emphasized in this paper is that it pays to invoke as many constraints as possible prior to determining the experimental design. This way, the domain of candidate designs is narrowed down to what has been dubbed a *reasonable design space*, and the approximation accuracy improves.[11,12] However, in this work odd designs are removed primarily in order to prevent ill-posed CFD problems, for which the diffuser performance is unsatisfactory and the analysis becomes more difficult. The introduction of wall monotonicity constraints has proven appropriate for the optimum shape design of diffusers.

II. Diffuser Design Case

Diffusers are commonly used to convert dynamic pressure into a static pressure rise by deceleration of a fluid flow. The problem of designing a diffuser for maximum pressure recovery has been found to be a well-suited touchstone for CFD optimization because of the relatively rapid and robust convergence conditions and the ease of interpreting optimum design solutions from underlying flow physics. Some prior examples of work on diffuser shape optimization include that performed by Çabuk and Modi, who used an inverse design method to design laminar diffusers,[13] and by Madsen, who used derivative-based optimization for turbulent flow diffusers.[14]

We introduce the dimensionless pressure recovery coefficient C_p as an objective function to be maximized:

$$F = C_p = \frac{\Delta p}{\frac{1}{2}\rho v_{\text{inl}}^2} \tag{1}$$

Here Δp is the static pressure difference between channel cross sections upstream and downstream of the diffuser respectively, ρ is the fluid density, and v_{inl} is the inlet mean velocity. Inlet and outlet static pressures are averaged, even though the pressure distribution is nearly uniform because of well-developed flow at the considered cross sections. The flow is incompressible and fully turbulent with a Reynolds number Re_D of 10^5, based on the inlet throat half-width D. The CFD model uses a symmetry condition along the channel center axis and has a computational mesh consisting of 120 × 50 cells including a long outlet section to establish a fully developed exit profile. The overall geometry of the two-dimensional planar diffuser (Fig. 1) is defined by the ratio of inlet and outlet areas AR and the diffuser length/height ratio L/D, where L is the axial length of the diffuser. In this study the L/D ratio is fixed at 3.0 and the area ratio AR at 2.0. Expressed in terms of the inlet half-width D, the horizontal length of the inlet section is $1D$, whereas the horizontal length of the outlet is $10D$.

The shape of the diffuser wall is designed for optimum performance, and to this end two separate cases of wall parameterizations are tried: 1) a two design variable case, where a polynomial describes wall shapes, and 2) a five design variable case that uses B-splines. Even though two different curve descriptions are used in the two cases, the most noteworthy difference seen from the point of view of the RSA lies in the problem size. Case 1 serves to give a good understanding of the methods, whereas case 2 is a more challenging optimization problem because of the larger number of design variables.

The two-dimensional flow analysis code applied is based on the full Reynolds-averaged Navier–Stokes equations, with the original k–ϵ two-equation turbulence model as a closure form.[15] For discretization second-order central difference schemes are applied to all derivatives except convection terms, which are treated by a second-order upwind scheme.[16,17] At the inlet of the flow domain, a uniform flow distribution is specified. At the outlet the first-order derivatives in the streamwise direction are set to zero for all dependent variables, except for the static pressure, which is regulated by an exit mass flux condition. At nodal positions next to solid walls, the wall function treatment is applied.[15] The sum of momentum and mass residuals, normalized by the inlet momentum and mass fluxes respectively, is used as solution convergence criterion. Computational meshes are curvilinear, structured, and generated using transfinite interpolation. Details of the CFD algorithm can be found in Ref. 16.

III. Design Parameterization

The paper treats the two cases of diffuser contour parameterization in parallel. In both, the design is controlled by the vertical positions of curve control points, which are uniformly spaced axially along the wall. Furthermore, it is required for both parameterizations that the wall shape is C^1-continuous at the inlet, in order to avoid a sharp edge leading to the onset of separation.

Two-Design-Variable Case

Because this case is mainly illustrative, a simple curve representation has been chosen, leading to the use of polynomial functions. Polynomials are often considered impractical for design optimization because of a tendency to produce undesirable wiggles. However, as will be shown here, for diffusers this problem can be eliminated by enforcing analytically derived conditions of curve monotonicity. The shape of the diffuser wall is described by a function $y(x)$, where y denotes the vertical position of wall points $y = 0$ at inlet and $y = 1$ at outlet, so that the inlet half-height D is the basis of a nondimensionalization. The independent variable x is the axial position along the curved wall section, with $0 \le x \le 10$, $x = 0$ at inlet and $x = 3$ at the end of the curved section, so that the curve connects endpoints with (x, y) coordinates (0, 0) and (3, 1) (Fig. 1). The wall contour in the curved section, for the two-design-variable case, is defined by a fourth-order polynomial:

$$y(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 \tag{2}$$

The vertical positions of two control points y_1 and y_2 are the design variables. The polynomial curve must pass through these control points, which are placed at horizontal positions $x = 1$ and 2. The use of control points as design variables yields stronger geometrical interpretation than the use of polynomial coefficients a_i. Given four points of the curve (two endpoints plus two control points) and the condition of zero slope $dy/dx = 0$ at the diffuser inlet ($x = 0$) to ensure C^1-continuity, the polynomial function reduces to

$$y(x) = a_4x^4 + a_3x^3 + a_2x^2 \tag{3}$$

The polynomial coefficients a_i are given in terms of control point positions y_1 and y_2 as follows:

$$a_2 = 3y_1 - 0.75y_2 + \tfrac{1}{9}, \qquad a_3 = -2.5y_1 + y_2 - \tfrac{1}{6}$$

$$a_4 = 0.5y_1 - 0.25y_2 + \tfrac{1}{18} \tag{4}$$

Five-Design-Variable Case

For shape parameterization in more variables, B-splines were preferred to natural splines (piecewise polynomials), although the latter technique is closer to the polynomial representation just described. B-splines excel in the predictable way that control points influence curve shape and in the local control, which prevents small changes in a control point position to propagate over the entire curve.[18] Combined with low computational cost, these advantages have contributed to B-spline curves becoming a standard geometric modeling technique in computer-aided design.

Unlike the explicit form of the polynomial curve, Eqs. (3) and (4), a B-spline is given in parametric form:

$$\begin{bmatrix} x(u) \\ y(u) \end{bmatrix} = \sum_{i=0}^{n} \boldsymbol{P}_i N_{i,k}(u) \tag{5}$$

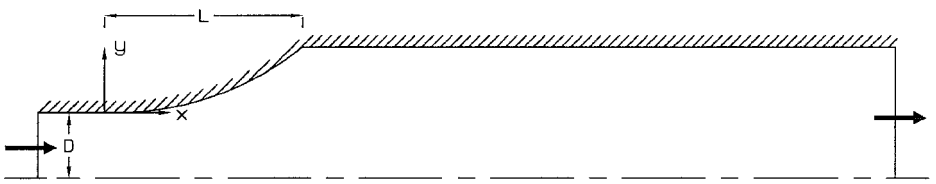

Fig. 1 Two-dimensional, symmetric diffuser subjected to shape optimization in terms of pressure recovery measured between inlet and outlet.

A set of blending functions $N_{i,k}$ combines the influence of $n + 1$ control points $\boldsymbol{P}_i$ over the range of the parametric variable u. The blending functions $N_{i,k}$ are recursively determined polynomials with degree $k - 1$, where the parameter k dictates the order of continuity of the curve and thus how many control points influence a curve segment.[18] In this work k is 8, which corresponds to C^6 continuity. The number of control points is eight as well: two endpoints, five design variables, and one point used for prescribing the inlet slope.

B-splines have an approximating nature in that they do not necessarily pass through control points, except for fixed-curve endpoints. The slope at a curve endpoint is tangential to a straight line connecting the endpoint and the first control point and can be prescribed by placing an additional fixed control point near the endpoint.

IV. Monotonicity Constraints

Given a limited number of analyzed designs from which to construct a response surface approximation, it becomes essential for good prediction accuracy to distribute these designs in a reduced size reasonable design space, in which designs a priori identified as bad have been removed. The constraints used to limit the domain may arise from any physical insight into the problem at hand. Notice that optimization via RSAs has the advantage that additional design constraints can easily be added, without need to rerun CFD analyses as is typically required when using search optimization. Experimental and numerical evidence indicates that maximum pressure recovery in diffusers occurs at the border of appreciable flow separation.[5,14,19] For this reason strongly separated diffuser flows should be avoided, which makes it reasonable to restrict the design space to monotonic wall shapes. Although the approximation accuracy does of course benefit from the reasonable design space approach, it is equally important in the present example that monotonicity constraints eliminate convergence problems associated with CFD analysis of odd, nonmonotonic designs.

Two-Design-Variable Case

Because the polynomial shape representation interpolates, i.e., the curve passes through the two control points, the monotonicity in wall shape was first replaced by requiring monotonicity in the position of control points:

$$0.0 \leq y_1 \leq y_2 \leq 1.0 \tag{6}$$

However, Eq. (6) is not sufficient to preclude all inappropriate designs with a nonmonotonic diffuser contour. From an analysis point of view, this is problematic because odd contours are allowed for which no sensible pressure recovery can be obtained.[8] Many experimental designs, including D-optimal ones, include a large fraction of points situated on edges of the design space. Such point selection is a sound strategy for reducing the effect of noise and for limiting the use of RSAs for extrapolation. However, edge points also challenge the robustness of numerical tools (grid generation and analysis) because they correspond to geometrically extreme designs. For effective use of RSA, the designer may need to develop criteria for precluding unreasonable designs. Equivalent problems occur less frequently in derivative-based search optimization, where move limit strategies prevent excessive design changes from iteration to iteration. Should, however, a search optimization venture into unreasonable configurations or cases with inadequate meshing, this is typically much more difficult to recognize than when using RSA, simply because erroneous results are more easily identified when they belong to a batch of analyses.

Because of the inadequacy of Eq. (6), strict conditions of monotonicity were derived and invoked. Madsen et al.[8] describe these derivations in detail. The strict monotonicity constraints reduce the design space to only about $\frac{1}{3}$ of the initial one given by Eq. (6). Both initial and reduced design spaces are shown in Fig. 2.

Five-Design-Variable Case

The parametric form in which B-splines are defined makes it nontrivial to derive monotonicity constraints analytically so that instead a constraint approximation $\hat{G}$ was set up in the form of a response surface for the minimum wall slope G. Then, observing the inequality constraint $G \geq 0$ implies a positive wall slope and thus monotonicity throughout. Because B-splines are inexpensive to generate, 9^5 (59,049) B-splines were computed (this took just a few seconds) and used for fitting a quadratic response surface. The approximation to the monotonicity constraint precludes some designs that satisfy the exact monotonicity requirement. However, the effect of these inaccuracies on the solution of the optimum design problem is negligible.[8]

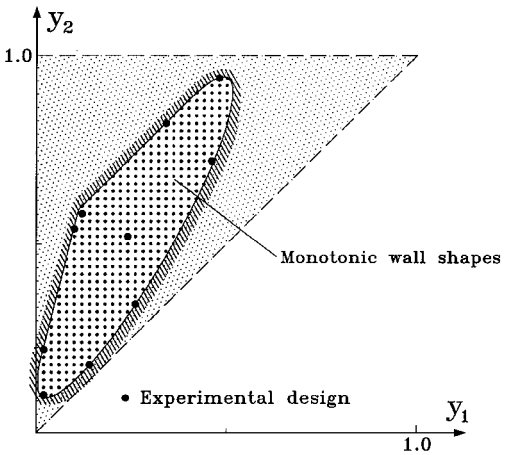

Fig. 2 Ellipse-like region is the monotonic portion of the design space, as found by applying strict monotonicity constraints. The initial design space is the shaded triangle bounded by – – –. · · · ·, Candidate points from which a D-optimal design is chosen. The chosen D-optimal set is indicated by •.

V. D-Optimal Experimental Design

In this work quadratic polynomials are used as RSAs. A quadratic polynomial (7) consists of $(n + 1)(n + 2)/2$ terms, i.e., six in the two-dimensional case and 21 in the five-dimensional one:

$$\hat{F} = b_0 + \sum_{i=1}^{n} b_i y_i + \sum_{i=1}^{n} \sum_{j=i}^{n} b_{ij} y_i y_j \tag{7}$$

Two-Design-Variable Case

The six coefficients of the two-dimensional quadratic polynomial were determined based on analyses at 10 sampling points $(y_1, y_2)_i$. These points were picked from a pool of approximately 400 designs, generated by discretizing each design variable in 50 levels and discarding designs violating the monotonicity constraints. From this pool a D-optimal subset was selected by means of the statistical software package JMP.[20]

D-optimality is among the best known and most often used criteria for point selection in designed experiments.[10] A D-optimal design minimizes the volume of the confidence region of the b_i coefficients and thus determines these with the highest degree of certainty. Figure 2 shows both the candidate set and the D-optimal design obtained. As it is typical for minimum-variance designs, the points chosen are concentrated around the edges of the design space. Notice in particular how two points in the experimental design are close neighbors from the candidate pool. In fact, the D-optimal design algorithm will choose a replicate design at one of these points if allowed. Such replication makes sense when the data come from experiments because it averages out experimental noise. For numerical experiments there is no sense in allowing exact replication, which will produce identical results, but the selection of close neighbors serves the corresponding purpose of averaging out numerical noise.

Five-Design-Variable Case

The regression analysis, to find 21 polynomial coefficients in five dimensions, is based on a 35-point D-optimal design. The surplus of analyses is generally required for reducing the sensitivity to numerical noise and to errors caused by the simplified representation as a quadratic polynomial. Again, a pool of candidate points was created, this time using nine levels for each variable (values ranging from 0.0 to 1.0) and then checking the monotonicity of the B-splines

for each of the $9^5 = 59{,}049$ designs. Observe that limiting the y coordinate of the control points to a variation in the range [0.0; 1.0] is a somewhat artificial requirement, as monotonic shapes exist with coordinates slightly outside this range. A total of 20,864 points are monotonic in wall shape. This relatively large percentage of acceptable cases reflects the smoother nature of approximating curves. Had a nonsegmented polynomial curve representation been used, the condition of monotonicity in the control points corresponding to Eq. (6) would alone have reduced the number of feasible design points to less than 1% of those inside a five-dimensional box. As in the two-design-variable case, the subset of D-optimal points was found using the JMP software.[20]

VI. Model Building

Whereas the response surface fit to a given set of data represents the best least-squares fit, it may not be the most accurate polynomial for predicting the function at other points. One source of predictive inaccuracies are polynomial coefficients that are poorly characterized by the data because they have little influence at the data points (but they can have large influence at other points). Another source of inaccuracies is outlier points having large errors caused by numerical noise, modeling errors or inadequacies of the software used to generate the data. JMP offers statistical tools for identifying and eliminating both sources of error.

A frequently used measure for the goodness of fit is the coefficient of multiple determination R^2, which measures the fraction of variation in data captured by the response surface. The remaining variation is attributed to random noise:

$$R^2 = SS_r/SS_y = 1 - SS_e/SS_y \tag{8}$$

Here SS_e is the sum of squared approximation errors at the n_p sampling points, SS_y is the true response's sum of squared variations from the mean $\bar{F}$, and similarly SS_r is the approximation's sum of squared variations from mean, where

$$SS_e = \sum_{i=1}^{n_p}(F_i - \hat{F}_i)^2, \qquad SS_y = \sum_{i=1}^{n_p}(F_i - \bar{F})^2$$

$$SS_r = SS_y - SS_e = \sum_{i=1}^{n_p}(\hat{F}_i - \bar{F})^2 \tag{9}$$

An R^2 value of one implies a perfect approximation, and values above 0.9 are normally needed for adequate approximation. The value of R^2 increases when terms are added to the approximation, reflecting a better fit to the given data. However, these additional terms may actually worsen the prediction accuracy of the response surface if they fit numerical noise (a problem called overfitting). Another measure R_a^2 (R-square adjusted) is better suited for assessing predictive accuracy. It is defined as

$$R_a^2 = 1 - \frac{SS_e/(n_p - n_\beta)}{SS_y/(n_p - 1)} \tag{10}$$

Here n_β is the number of regression coefficients. Like R^2, the values of R_a^2 range between zero and one, and the higher values indicate better predictive accuracy. The value of R_a^2 will often decrease when unnecessary terms are added to the model. Moreover, a large difference between the values of R^2 and R_a^2 is a good indication of insignificant model terms.[10]

Two-Design-Variable Case

For the parameterization in two design variables, the quadratic response surface includes six terms, where

$$\hat{F} = b_0 + b_1 y_1 + b_2 y_2 + b_3 y_1^2 + b_4 y_1 y_2 + b_5 y_2^2 \tag{11}$$

Here, a backward elimination procedure based on t statistics is used to discard terms and improve the prediction accuracy.[10] The t statistic of a fitting coefficient is its value divided by an estimate of the standard error of the coefficient. A common rule of thumb is that regressor terms with an absolute t-statistic value larger than two are significant at a 95% confidence level. The backward elimination procedure consists of removing the most uncertain regressor, i.e., the coefficient with the smallest absolute value of the t statistic, and performing another regression with the remaining terms, carrying on as long as an improvement in prediction can be obtained.

Table 1 Regressor values and t statistics from linear regression analysis

	Full		Reduced	
Coefficient	Value	t statistic	Value	t statistic
b_0	0.6608	153.05	0.6615	185.82
b_1	0.2192	5.49	0.2283	7.84
b_2	0.0688	1.95	0.0584	2.85
b_3	−0.3001	−1.78	−0.2408	−4.10
b_4	0.0697	0.38	——	——
b_5	−0.1039	−1.46	−0.0783	−3.67

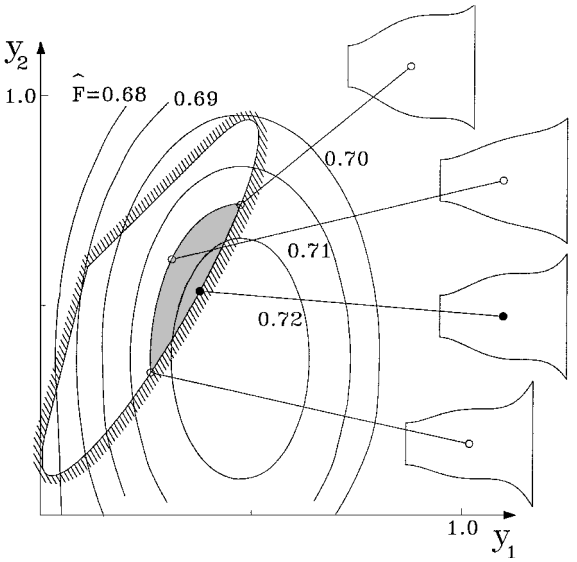

Fig. 3 Contour plot of the response surface values $\hat{F}$ with ● indicating the optimum design point. The shaded part of the feasible design space comprises designs with performance within 1% of the optimal. Some of the corresponding shapes are indicated to the right.

Table 1 lists the regression coefficients b_i and their t statistics, for both the complete quadratic polynomial (11) and a partial model arising from the backward elimination procedure.

The full response surface, including all six terms, has an R_a^2 of 0.956. The t statistics in Table 1 suggest that the determination of the mixed-term regressor b_4 involves a large standard error, estimated to about three times the coefficient itself. Discarding this term, the new regression analysis with a partial model of five terms raised R_a^2 to 0.964. Meanwhile, the R^2 value decreased almost imperceptibly from 0.981 to 0.980. The relatively high values of R^2 and R_a^2 seem to indicate a very good fit. According to Table 1, removal of the mixed-term regressor also improved the standard errors of the remaining terms substantially so that these all become significant at the 95% confidence level. Further elimination of terms did not improve the fit so that the best quadratic RSA of the diffuser pressure recovery is the following:

$$\hat{F} = 0.6615 + 0.2283 y_1 + 0.0584 y_2 - 0.2408 y_1^2 - 0.0783 y_2^2 \tag{12}$$

A contour plot of this response surface along with the posed monotonicity constraints is shown in Fig. 3.

To check predictive accuracy, four points are used for a comparison between CFD results, the full response surface, and the preferred partial model. Results are listed in Table 2, and as seen the differences between the two surfaces are small, but the reduced surface is more accurate for three out of the four points. Furthermore, there is a good agreement between numerical experiments and approximations.

Table 2 Comparison between CFD solutions and RSA predictions

(y_1, y_2)	F	$\hat{F}$ (full)	$\hat{F}$ (reduced)
(0.1, 0.3)	0.6968	0.6931	0.6924
(0.3, 0.5)	0.7169	0.7184	0.7179
(0.35, 0.55)	0.7182	0.7206	0.7203
(0.4, 0.65)	0.7176	0.7194	0.7191

Table 3 Backward elimination procedure for RSA in five variables

Terms	Min $\|t\|$	No. $\|t\| < 2.0$	R^2	R_a^2	Comments
21	0.05	15	0.922	0.811	——
20	0.23	14	0.922	0.823	Removed y_3^2 term
19	0.45	12	0.922	0.834	Removed y_4^2 term
18	0.53	9	0.921	0.841	Removed $y_1 y_4$ term
17	0.97	8	0.919	0.848	Removed $y_2 y_5$ term
*16	1.22	6	0.915	0.848	Removed y_2^2 term
15	1.57	5	0.909	0.844	Removed $y_1 y_3$ term

Table 4 Comparison between CFD solutions and RSA predictions

y_1	y_2	y_3	y_4	y_5	F	$\hat{F}$ (full)	$\hat{F}$ (reduced)
0.5	0.5	0.5	0.5	0.5	0.7171	0.7148	0.7126
1.0	0.5	0.0	0.5	1.0	0.7174	0.7210	0.7174
0.25	0.75	0.25	0.75	0.25	0.7148	0.7185	0.7162
0.0	0.5	1.0	0.5	0.0	0.6943	0.7333	0.7283

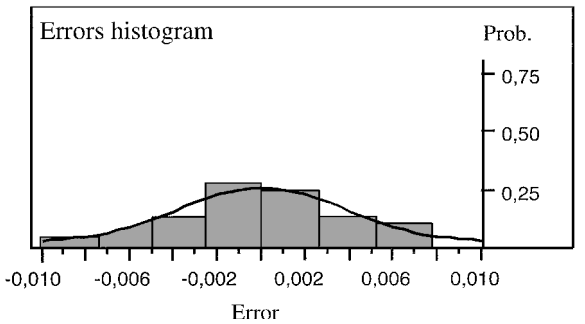

Fig. 4 Distribution of response surface errors at sampling points and the corresponding normal distribution curve (same mean and variance).

Five-Design-Variable Case

Based on the D-optimal set of 35 design points selected, the 21 regressors of a full quadratic polynomial were fitted resulting in a moderate R_a^2 value of 0.810. A backward elimination of regressor terms subsequently led to the removal of five terms and an increase of R_a^2 to 0.848. The lower values of R_a^2, in comparison to the two-design-variable case, reflect the increased difficulties in obtaining a good fit when moving to higher-dimensional response surfaces. Data on the backward elimination steps are given in Table 3, which apart from R^2 and R_a^2 hold the minimum t statistic and the number of uncertain terms with $|t| < 2.0$ remaining in the model. From the t-statistics information it appears that the backward elimination improved the accuracy of remaining terms.

The next step performed was to investigate whether the 35 applied observations included outliers. A common (but not necessarily true) assumption, which enables the statistical treatment of observations, is that errors are independently and identically distributed according to a normal distribution with mean zero and variance. Thus, the distribution of response surface errors was plotted and compared to a normal distribution, with which it is expected to correspond well. From the histogram plot of the error distribution (see Fig. 4), it did not seem that there are any outliers.

As in the two-design-variable case, four arbitrary points were picked to test the prediction accuracy of the RSA away from sampling points. Table 4 compares CFD results and polynomial approximations with and without backward elimination of terms.

Again, the predictions of the response surface appear reliable, apart from at the last control point. This point is, however, well in the nonmonotonic region so that the approximation relies on an extrapolation, which was never intended. The reduced approximation model comes closer to the CFD results for two out of the three meaningful test points.

VII. Numerical Noise

Although noisy data from laboratory experiments is a generally accepted fact, the presence of noise in numerical simulations seems much less recognized. Because of the complex numerical modeling techniques of CFD, the exact origins of noisy responses may be difficult to pinpoint, but factors such as turbulence models, incomplete convergence, and the discretization itself are certainly influential. Here, the presence of numerical noise will be investigated. The problem of nonsmooth or noisy objective functions has previously been addressed by Giunta et al., who found RSA-based optimization to perform very robustly under such circumstances, especially when point selection is based on design of experiment techniques, such as D-optimal designs.[21]

The noise in the objective function was investigated in the following way for the two-design-variable case. First, two arbitrary (feasible) designs, (0.3, 0.6) and (0.31, 0.61), were chosen, and nine designs were distributed uniformly along the straight line connecting these two design points. Then, CFD computations were carried out for each of the 11 points. The variation in performance along this arbitrary line in design space is plotted in Fig. 5. The noticeable oscillation of objective function values is of such a magnitude that it must be triggered by other sources than simply bad conditioning. The nonsmooth behavior of flow response could create difficulties in the use of gradient-based optimization algorithms, even though no practical evidence thereof exists from the study of similar problems.[5,14] For comparison, the constructed response surface, Eq. (12), is also plotted in Fig. 5, which neatly illustrates the smoothing effect of these approximations.

It is tempting to dismiss the noisy responses on the ground of modeling errors in the CFD code applied, but this would be a rash conclusion. As the magnitude of noise was so unexpectedly high (around 1.0%), the same cases were solved using another finite volume based CFD code.[22] The capabilitites and methods of this other CFD code are essentially identical to those of the code used throughout, besides the fact that it operates with a collocated instead of staggered variable arrangement. Furthermore, because of limitations of the software used, a different wall-shape parameterization was applied in this second investigation of noise. A B-spline curve with two free control points was used instead of a fourth-order polynomial function. Again, the observation was made that the objective function oscillated because of numerical noise, but the amplitude was much smaller than in Fig. 5. To make the noise more apparent, it was therefore necessary to refine the subdivision of the discretized line and reduce its length to 20% of the initial so that the line spans from (0.3, 0.6) to (0.302, 0.602).

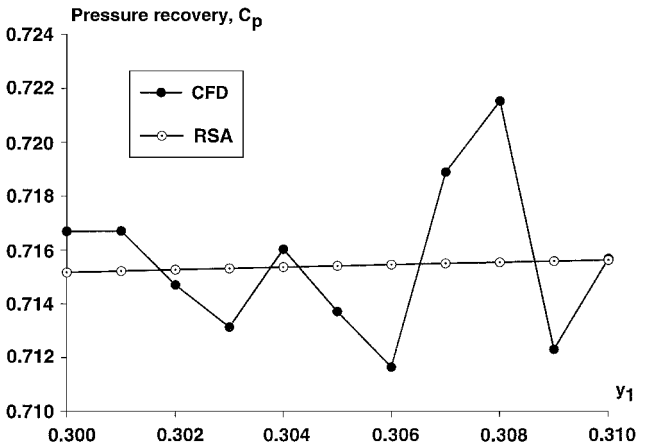

Fig. 5 C_p values from analyses at 11 design points along an arbitrarily chosen straight line in the design space.

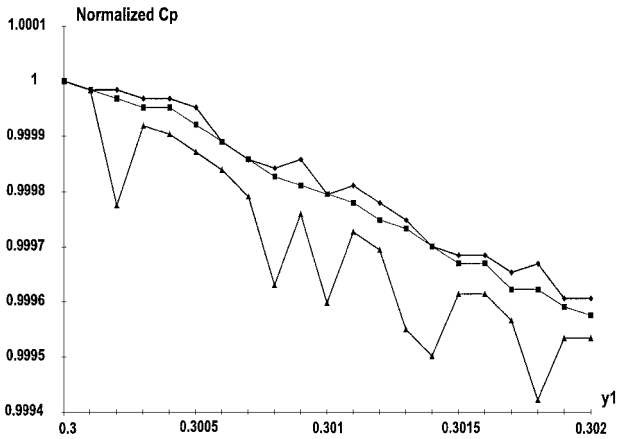

Fig. 6 Normalized C_p values along straight line in design space. The results are for two different differencing schemes and two different residual levels used as convergence criterion: ■, UDS; ◆, SUDS (tight tolerance); and ▲, SUDS (loose tolerance).

This yielded the noisy response patterns shown in Fig. 6. The two topmost curves were determined using a relatively tight convergence criterion and two different convection schemes—a standard first-order upwind differencing scheme (UDS) and a second-order upwind differencing scheme (SUDS). The use of different differencing schemes was carried out to estimate whether numerical diffusion does significantly dampen the generation of noise. Comparing the two codes applied, there are indications that the one used throughout is less forgiving, in the sense that it predicts a stronger tendency for flow separation. This could possibly be explained by factors such as numerical diffusion, boundary treatments, and momentum interpolation methods adopted in the two codes. As expected, switching to a more dissipating differencing scheme (lower-order accuracy) yields a smoother response. To further illustrate this issue, one more design line curve is shown in Fig. 6, which arose from using a relatively loose, yet still reasonable, convergence criterion (using SUDS). The applied convergence criterion considers summed and normalized (by inlet flux) residuals over the entire mesh, with termination of computations once the maximum is below a certain small value ϵ. The loose convergence criterion in Fig. 6 was $\epsilon = 10^{-3}$, whereas the tight tolerance was $\epsilon = 10^{-5}$. For comparison, a convergence limit of $\epsilon = 10^{-4}$ was applied in the CFD analyses used for response surface modeling. The overall conclusion is that the presence of some numerical noise in CFD results is practically inevitable, although its magnitude depends on choice of code and modeling techniques. Here, a technique such as RSA can be effective in smoothing out the undesirable fluctuations.

VIII. Optimum Diffuser Designs

Two-Design-Variable Case

Based on the constructed response surface approximation, Eq. (12), and the monotonicity constraints, the optimum design was found at $y_1^* = 0.3767$ and $y_2^* = 0.5333$, which is indicated by a solid circle in Fig. 3. The predicted optimum value of C_p was 0.7220 and for comparison a CFD analysis yielded 0.7185. This is slightly off, but in any case there is a wide range of geometries that yield essentially indistinguishable performance. Figure 3 shows a 1% tolerance region around the optimum design and also gives some examples of geometries within the range yielding similar performances. Moreover, the optimum detected from a full quadratic response surface model ($y_1^* = 0.3775$ and $y_2^* = 0.5348$), including all six terms, is almost identical with the optimum just stated. The appearance of the sharp corner in Fig. 3, evidenced in all diffuser shapes shown there, results from the geometric constraint selected in the present study. Although in reality one will most likely avoid creating such a feature, its presence makes insignificant impact on pressure recovery compared to modeling accuracy.

For two-dimensional diffusers without flow separation, the demonstration has been made through boundary-layer calculations and experiments[23,24] that the highest performance is to be expected from a convex-outward or bell-shaped design. This can be explained by the main cross-sectional enlargement ideally happening in the start of the diffuser, where the boundary layer is still stable and able to withstand a higher adverse pressure gradient. In accordance with these findings, the optimum-designed diffuser is mostly bell shaped. However, the end of the diffuser wall is bent concavely outward. In this part of the diffuser, the flow separates, which seems inevitable given the area expansion ratio and overall geometry of the diffuser. The very nice bell shape before the inflection point indicates that optimization has cleverly adjusted the wall slope to delay the onset of separation as far as possible because the presence of separation accounts for large losses.

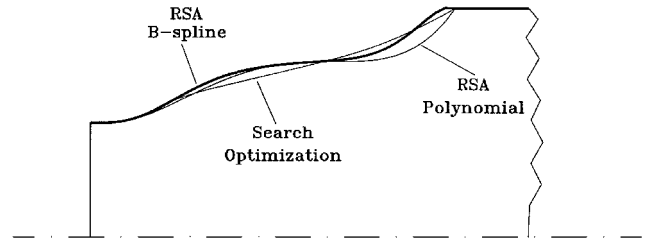

Fig. 7 Comparison of optimum wall shapes using polynomial and B-spline representations, respectively.

Five-Design-Variable Case

In the optimum design using B-spline parameterization, both the monotonicity constraint and four out of five side constraints are active. As already mentioned, the response surface constructed to guarantee wall monotonicity becomes too restrictive. To compensate for this, a one-dimensional search in the direction of the steepest gradient was conducted starting at the optimum design point estimated by RSA:

$$y = y^* + \alpha \nabla F \tag{13}$$

This search is terminated as soon as designs turn nonmonotonic, yielding a new optimum point at the edge of the true feasible domain and an increase in the optimum pressure recovery coefficient from 0.7208 to 0.7235.

Figure 7 compares the optimum wall contours determined by RSA using B-splines and polynomial shapes. The optimum B-spline shape compares well to the optimum polynomial one, and so it is not surprising that there is no significant gain compared to this case. The largest differences in shape are found in the later part of the expansion, where the shape has less impact on the overall performance, as separation is small in either case. Thus, the close resemblance of optimum inlet shapes is reassuring in terms of the credibility of the optimization algorithm. A CFD analysis of the five-design-variable optimum design yields a pressure recovery coefficient of 0.7193, a little below the predicted value, as in the two-design-variable case. The improvement from the two-design-variable case (0.7185–0.7193) indicates that there is not much potential for further gains. Furthermore, for comparison, Fig. 7 also contains the corresponding wall contour determined using gradient-based search optimization techniques.[14] The optimum wall shape found by search optimization can be described as truly bell shaped, without a plateau similar to the one found in the results of RSA optimization. There appears to be a distinct difference in optimum shapes from the two different optimization approaches, which must be ascribed to the combination of optimization accentuating modeling differences and a relatively small scatter in diffuser performances.

IX. Conclusion

The present work illustrates the use of response surface approximations for the solution of optimum design problems in fluid mechanics. The wall shape of a two-dimensional diffuser was optimized in order to achieve maximum pressure recovery, with two different curve representations for the wall.

Response surface techniques provide an easy way to connect a black-box CFD code to an optimizer. In addition, response surfaces

are useful to filter out numerical noise, which is often generated by CFD codes and which may complicate optimization. The magnitude of noise is influenced by the specific modeling techniques applied, but because of their complexity, it is difficult to pinpoint the cause of the noise in order to eliminate it. The grid size used in the present work is quite coarse based on today's standards. However, in design practice the use of coarse grid solutions is attractive because of the economy. In the present work insight into noise issues was obtained by comparing the present analysis to another CFD code.

Response surface modeling based on analysis points selected by design of experiments was found to challenge the robustness of analysis codes by leading to the envelope of permissible geometrical configurations, thus creating potential difficulties in convergence. In derivative-based search optimization with a reasonable move limit strategy, this rarely happens. It was necessary to implement sophisticated monotonicity constraints to circumvent some of these problems. The restricted design space, containing diffuser designs with monotonic wall shapes only, was reduced to about $\frac{1}{3}$ of the initial domain. This efficiently excluded ill-posed analysis problems and furthermore improved approximation accuracy.

Response surface techniques include a variety of statistical tools aimed at the elimination of poorly characterized model effects. However, these tools are more in the nature of safety measures, and often they do not show any payoff. In our case two ways of improving the fit were considered, namely a method to eliminate uncertain regressor terms and a technique to spot and eliminate outlier design points (i.e., runs where something went wrong). The benefit from eliminating coefficients was insignificant, possibly because the danger from including poorly determined coefficents is more pronounced in higher-dimensional cases, where extrapolation is inevitable. Also, no outliers were found, and the distribution of errors agreed well with the assumed statistical model of normal distribution.

The response surface optimization resulted in a diffuser design, which is mostly bell shaped with the end of the diffuser wall bent concavely outward. This design was compared to another solution of the same problem, found with a different CFD code and a derivative-based optimizer, and the optimal shapes obtained were quite different. The tendency of optimization to accentuate differences between models has been amplified in the present problem by the relative insensitivity in diffuser performance to shape. Thus, when comparing the two CFD codes, the differences in the analysis of most configurations is small, but the difference between the final design shapes is striking. The gain in pressure recovery from optimization is small, as there is a large family of diffusers with essentially identical performance. For the present case, with the physical/numerical confidence of our model, it is impossible to tell better than 1%, which means that within the confidence level of the CFD tool the satisfactory design is not unique. In conclusion, one should watch for situations where the optimizer tends to place the optimum on the brink of infeasibility.

Acknowledgments

The work of the first author was supported by the Danish Technical Research Council. The work of the third author was supported by NASA Grant NAG1-2000.

References

[1]Baysal, O., and Eleshaky, M. E., "Aerodynamic Design Optimization Using Sensitivity Analysis and CFD," *AIAA Journal*, Vol. 30, No. 3, 1992, pp. 718–725.

[2]Lambert, P. A., Lecordix, J. L., and Braibant, V., "Constrained Optimization of Nacelle Shapes in Euler Flow Using Semi-Analytical Sensitivity Analysis," *Structural Optimization*, Vol. 10, No. 3/4, 1995, pp. 239–246.

[3]Elbanna, H. M., and Carlson, L. A., "Aerodynamic Sensitivity Coefficients Using the Three-Dimensional Full Potential Equations," *Journal of Aircraft*, Vol. 31, No. 5, 1994, pp. 1071–1077.

[4]Taylor, A. C., Hou, G. W., and Korivi, V. M., "Methodology for Calculating Aerodynamic Sensitivity Derivatives," *AIAA Journal*, Vol. 30, No. 10, 1992, pp. 2411–2419.

[5]Svenningsen, K. H., Madsen, J. I., Páuker, W. H. G., and Hassing, N. H., "Optimization of Flow Geometries Applying Quasi-Analytical Sensitivity Analysis," *Applied Mathematical Modelling*, Vol. 20, No. 3, 1996, pp. 214–224.

[6]Sobieszczanski-Sobieski, J., and Haftka, R. T., "Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments," *Structural Optimization*, Vol. 14, No. 1, 1997, pp. 1–23.

[7]Narducci, R., Grossman, B., Valorani, M., Dadone, A., and Haftka, R. T., "Optimization Methods for Non-Smooth or Noisy Objective Functions in Fluid Design Problems," AIAA Paper 95-1648, June 1985.

[8]Madsen, J. I., Shyy, W., Haftka, R. T., and Liu, J., "Response Surface Techniques for Diffuser Shape Optimization," AIAA Paper 97-1801, June 1997.

[9]Johansen, M. D., Lauridsen, S., Risager, L., Madsen, J. I., and Lundgren, K. D., "Response Surface Approximations for Shape Optimization of a 3D-Manifold," *Proceedings of 4th CFX International Users Conference*, AEA Technology, Chicago, 1997, pp. 222–231.

[10]Myers, R. H., and Montgomery, D. C., *Response Surface Methodology—Process and Product Optimization Using Designed Experiments, Series in Probability and Statistics*, Wiley, New York, 1995.

[11]Kaufman, M., Balabanov, V., Burgee, S. L., Giunta, A. A., Grossman, B., Haftka, R. T., Mason, W. H., and Watson, L. T., "Variable-Complexity Response Surface Approximations for Wing Structural Weight in HSCT Design," *Computational Mechanics*, Vol. 18, No. 2, 1996, pp. 112–126.

[12]Roux, W. J., Stander, N., and Haftka, R. T., "Response Surface Approximations for Structural Optimization," *International Journal for Numerical Methods in Engineering*, Vol. 42, No. 3, 1998, pp. 517–534.

[13]Çabuk, H., and Modi, V., "Optimum Plane Diffusers in Laminar Flow," *Journal of Fluid Mechanics*, Vol. 237, 1992, pp. 373–393.

[14]Madsen, J. I., "Design Optimization of Internal Flow Devices," Ph.D. Dissertation, Inst. of Energy Technology and Inst. of Mechanical Engineering, Aalborg Univ., Aalborg, Denmark, Oct. 1998.

[15]Launder, B. E., and Spalding, D. B., "The Numerical Computation of Turbulent Flows," *Computer Methods in Applied Mechanics and Engineering*, Vol. 3, No. 2, 1974, pp. 269–289.

[16]Shyy, W., *Computational Modeling for Fluid Flow and Interfacial Transport*, Elsevier, Amsterdam, 1994, pp. 113–163.

[17]Thakur, S., Wright, J., Shyy, W., Liu, J., Ouyang, H., and Vu, T., "Development of Pressure-Based Composite Multigrid Methods for Complex Fluid Flows," *Progress in Aerospace Science*, Vol. 32, No. 4, 1996, pp. 313–375.

[18]Mortenson, M. E., *Geometric Modeling*, Wiley, New York, 1985, pp. 30–150.

[19]Kline, S. J., Abbott, D. E., and Fox, R. W., "Optimum Design of Straight-Walled Diffusers," *Journal of Basic Engineering*, Vol. 81, 1959, pp. 321–329.

[20]*JMP Users Guide*, Version 3.1, SAS Institute, Inc., Cary, NC, 1995.

[21]Giunta, A. A., Dudley, J. M., Narducci, R., Grossman, B., Haftka, R. T., Mason, W. H., and Watson, L. T., "Noisy Aerodynamic Response and Smooth Approximations in HSCT Design," AIAA Paper 94-4376, Sept. 1994.

[22]Sørensen, N. N., "General Purpose Flow Solver Applied to Flow over Hills," Ph.D. Dissertation, Risø National Lab., Risø-R-827 (EN), Roskilde, Denmark, June 1995.

[23]Reneau, L. R., Johnston, J. P., and Kline, S. J., "Performance and Design of Straight, Two-Dimensional Diffusers," *Journal of Basic Engineering*, Vol. 89, March 1967, pp. 141–150.

[24]Carlson, J. J., Johnston, J. P., and Sagi, C. J., "Effects of Wall Shape on Flow Regimes and Performance in Straight, Two-Dimensional Diffusers," *Journal of Basic Engineering*, Vol. 89, March 1967, pp. 151–160.

A. Messac
Associate Editor