Design and Evaluation of Homotopies for Efficient and Robust Continuation

David A. Brown and David W. Zingg

University of Toronto Institute for Aerospace Studies, Toronto, Ontario, M3H 5T6, Canada

Abstract

Homotopy continuation, in combination with a quasi-Newton method, can be an efficient and robust technique for solving large sparse systems of nonlinear equations. The homotopy itself is pivotal in determining the efficiency and robustness of the continuation algorithm. As the homotopy is defined implicitly by a nonlinear system of equations to which the analytical solution is by assumption unknown, many properties of the homotopy can only be studied using numerical methods. The properties of a given homotopy which have the greatest impact on the corresponding continuation algorithm are traceability and linear solver performance. Metrics are presented for the analysis and characterization of these properties. Several homotopies are presented and studied using these metrics in the context of a parallel implicit three-dimensional Newton-Krylov-Schur flow solver for computational fluid dynamics. Several geometries, grids, and flow types are investigated in the study. Additional studies include the impact of grid refinement and the application of a coordinate transformation to the homotopy as measured through the traceability and linear solver performance metrics.

Keywords: homotopy, continuation, computational fluid dynamics, high-dimensional curves, implicitly-defined curves, curvature, traceability

1. Introduction

Consider a curve segment defined implicitly by the system of equations

$$\mathcal{H}\left(\mathbf{q}\left(\lambda\right),\lambda\right)=\mathbf{0}, \tag{1}$$

$\mathcal{H}:\mathbb{R}^N\times\mathbb{R}\rightarrow\mathbb{R}^N$, $\mathbf{q}\in\mathbb{R}^N$, $\lambda\in\mathbb{R}$ on some interval $\lambda\in\Lambda$, $\Lambda\subset\mathbb{R}$. Without loss of generality, let $\Lambda=[0,1]$. In this paper we assume that $\mathcal{H}$ is at least C^1 differentiable, invertible, and that the curve is regular on Λ. As such, the curve derivatives cannot vanish and no bifurcations are present. While bifurcation analysis has been of practical interest to many researchers, including several in computational fluid dynamics [33, 36, 44, 45], the application of interest in this paper is the design of efficient homotopy continuation algorithms, for which the construction of bifurcated curves should be avoided as such curves are difficult and computationally expensive to trace numerically.

The curve defined implicitly by equation (1) can also be interpreted as a deformation. If the deformation is continuous, it is called a homotopy [1]. In the case where $\mathcal{H}\left(\mathbf{q},1\right)=\mathbf{0}$ is easy to solve numerically and $\mathcal{H}\left(\mathbf{q},0\right)=\mathbf{0}$ is difficult to solve numerically, solving $\mathcal{H}\left(\mathbf{q},1\right)=\mathbf{0}$

for $\mathbf{q}$ and approximately tracing the curve numerically from $\lambda = 1$ to $\lambda = 0$ can be an efficient and robust strategy [4, 6, 13, 14, 16, 43, 46] for acquiring an approximation to the solution to $\mathcal{H}(\mathbf{q}, 0) = \mathbf{0}$, where $\mathcal{R}(\mathbf{q}) = \mathcal{H}(\mathbf{q}, 0)$ is a system of equations for which the solution is of interest. This is referred to as homotopy continuation [1].

The use of homotopy continuation for its efficiency and robustness is a recent research area in computational fluid dynamics. Part of the attraction of homotopy continuation for large-scale scientific computing is the performance scalability as the number of degrees of freedom in the calculation increases. This is particularly attractive for future applications as the average problem size continues to increase, precipitated by the increasing availability and power of computational hardware. Performance scaling with grid refinement or discretization order has been investigated numerically by several authors [6, 13, 46] and is further investigated in this paper.

There are unlimited different ways in which to construct a regular homotopy $\mathcal{H}(\mathbf{q}, \lambda) = \mathbf{0}$ satisfying $\mathcal{H}(\mathbf{q}, 0) = \mathcal{R}(\mathbf{q})$ for a given $\mathcal{R}(\mathbf{q})$. For our application of interest, the homotopy should be constructed to maximize the efficiency and robustness of the curve-tracing algorithm. Homotopies targeting this application have so far been designed with little to ensure that the curve will be easy to trace or result in an efficient algorithm.

Some studies have been performed of homotopies in the context of efficient continuation. Hicken and Zingg [16] calculated eigenvalues along the continuation path to give an idea of the performance of the nonlinear and linear system solvers as a function of λ. For a simple one-dimensional problem, Hao *et al.* [13] plotted the solution field at several values of λ to give some intuitive visualization of the homotopy. For more complex three-dimensional systems, Brown and Zingg [6] tracked functionals as surrogates for the homotopy curve. However, even combining all of these approaches gives an incomplete profile of the homotopy and important information pertinent to continuation algorithm performance is still lacking.

Timing comparisons are not performed in this paper as they have been performed previously by Brown and Zingg [6], who found performance to be competitive or superior to the popular pseudo-transient continuation algorithm over an extensive suite of computational aerodynamics test cases. The focus of this paper is on identifying and quantifying features of homotopies which affect the performance of the continuation algorithm and in improving our understanding of why some homotopies lead to better algorithm performance. This knowledge will improve our ability to design effective homotopy continuation algorithms. The methodology is demonstrated by considering some candidate homotopies for the external aerodynamic flow solver of Hicken and Zingg [15] and Osusky and Zingg [29].

2. Homotopy

2.1. Homotopy Continuation

Consider the so-called convex homotopy [1] which is defined as the (presumably) continuous solution $\mathbf{q}(\lambda)$ to

$$\mathcal{H}(\mathbf{q}, \lambda) = (1 - \lambda)\mathcal{R}(\mathbf{q}) + \lambda\mathcal{G}(\mathbf{q}) = \mathbf{0}, \tag{2}$$

$$\mathcal{H} : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N, \ \mathcal{G} : \mathbb{R}^N \to \mathbb{R}^N, \ \mathcal{R} : \mathbb{R}^N \to \mathbb{R}^N, \ \lambda \in \mathbb{R}.$$

A continuation method can be developed from this homotopy by discretizing in λ to form a sequence of nonlinear equations:

$$\mathcal{H}(\mathbf{q}, \lambda_k) = (1 - \lambda_k)\mathcal{R}(\mathbf{q}) + \lambda_k\mathcal{G}(\mathbf{q}) = \mathbf{0}, \tag{3}$$

$$k \in [0, m],\ \lambda_k \in \mathbb{R},\ \lambda_0 = 1,\ \lambda_m = 0,\ \lambda_{k+1} < \lambda_k.$$

In the context of the studies in this paper, $\mathcal{R}(\mathbf{q})$ is the discrete flow residual and $\mathcal{G}(\mathbf{q})$ a system of equations of our own design which we refer to as the homotopy system.

Another form of homotopy continuation, known as global or Newton homotopy continuation due to its original formulation as a globally convergent generalization of Newton's method [1, 7, 26], is performed by sequentially solving

$$\mathcal{H}(\mathbf{q}, \lambda_k) = \mathcal{R}(\mathbf{q}) - \lambda_k \mathcal{R}(\mathbf{q}_0) = \mathbf{0}, \tag{4}$$

$$k \in [0, m],\ \lambda_k \in \mathbb{R},\ \lambda_0 = 1,\ \lambda_m = 0,\ \lambda_{k+1} < \lambda_k,$$

$$\mathcal{H} : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N,\ \mathcal{R} : \mathbb{R}^N \to \mathbb{R}^N.$$

The vector $\mathbf{q}_0 \in \mathbb{R}^N$ can be any vector of choice; for the computational aerodynamics examples presented in this paper, this vector is populated with the far-field boundary conditions.

2.2. *Predictor-Corrector Algorithm for Homotopy Continuation*

As the name suggests, the predictor-corrector algorithm consists of two phases: the predictor phase and the corrector phase. The two phases are applied repeatedly until traversing is complete.

The objective of the predictor phase is to obtain a suitable starting guess for the $k+1$st sub-problem using the estimated solution at the k-th sub-problem, a trajectory, and a distance (step-length) to travel along that trajectory. A common choice of predictor is the Euler predictor, for which the update at the kth step is given by

$$\mathbf{u}_{k+1}^{(0)} = \mathbf{u}_k^{(p_k)} + h_k \mathbf{d}_k, \tag{5}$$

$$h_k \in \mathbb{R}_+,\ \mathbf{u}_k \in \mathbb{R}^{N+1},\ \mathbf{d}_k \in \mathbb{R}^{N+1},$$

where $\mathbf{u}_k = [\mathbf{q}_k; \lambda_k]$, h_k is the step-length, $\mathbf{d}_k$ is the step direction, and $p_k \in \mathbb{Z}$ (where $\mathbb{Z}$ is the set of integers) is the number of iterations required to converge the k-th sub-problem. The step direction d_k is typically an estimate of the vector tangent to the curve. The step-length h_k is typically initialized at the first step and adjusted throughout the traversing process using a step-length adaptation algorithm [1, 4].

A variety of higher-order predictors do exist and have been shown to be effective for certain cases. Higher-order predictors can be based, for example, on polynomial extrapolation [1], Adams-Bashforth [24], or Runger-Kutta [2] methods. However, all examples of which we are aware of homotopy continuation algorithms applied to computational fluid dynamics problems have used Euler predictors, most likely due to their relative simplicity.

The objective of the corrector phase is to solve the nonlinear sub-problem at λ_k. The sub-problem $\mathcal{H}(\mathbf{q}_k, \lambda_k) = \mathbf{0}$ is solved inexactly using the inexact Newton method. Newton iterations are performed until the relative residual $\left\|\mathcal{H}\left(\mathbf{q}_k^{(n)}, \lambda_k\right)\right\| / \left\|\mathcal{H}\left(\mathbf{q}_k^{(0)}, \lambda_k\right)\right\|$ is reduced below some user-defined tolerance $\mu_k \in \mathbb{R}$.

2.3. *Parametrization*

The most intuitive way to interpret the homotopy curve is to consider the parametric curve $\mathbf{q} \in \left\{\mathcal{H}^{-1}(\mathbf{0}) \,|\, \lambda \in [0, 1]\right\}$, $\mathbf{q} \in \mathbb{R}^N$, $(\lambda) \mapsto \mathbf{q}(\lambda)$. This homotopy curve will be referred to as having λ parametrization. However, the curve can be written in a more general parametrized form $c(s) \in \left\{\mathcal{H}^{-1}(\mathbf{0}, \lambda) \,|\, \lambda \in [0, 1]\right\}$, $(s) \mapsto c(s)$, $c : \mathbb{R} \to \mathbb{R}^N \times \mathbb{R}$, where $c(s)$ is of the form $(\mathbf{q}(s); \lambda(s))$ and $s \in \mathbb{S}$, $\mathbb{S} \subset \mathbb{R}$. It is common in homotopy curve tracing algorithms to use an arclength parametrization as a reference frame [1, 4, 6].

Definition 1. *Let $c : \mathbb{R} \to \mathbb{R}^M$, $(s) \mapsto c(s)$ be a C^1-differentiable curve in $\mathbb{R}^M$ with parameter $s \in \mathbb{S}$, $\mathbb{S} \subset \mathbb{R}$. The parametrization defined implicitly by*

$$\dot{c}(s) \cdot \dot{c}(s) = 1 \tag{6}$$

is called an arclength parametrization of $c(s)$ [22].

A formal definition of the λ parametrization can also be given in the following form.

Definition 2. *Let $c : \mathbb{R} :\to \mathbb{R}^M \times \mathbb{R}$, $(r) \mapsto c(r)$ be a C^1-differentiable curve in $\mathbb{R}^M \times \mathbb{R}$ of the form $(\mathbf{q}(r); \lambda(r))$ with parameter $r \in \mathbb{R}$, $r \in [0, 1]$. The parametrization defined by*

$$\dot{\lambda}(r) = -1 \tag{7}$$

is called a λ parametrization of $c(r)$.

From this point forward, the notation $c(s)$ is used to indicate that the curve has an arclength parametrization, while $c(r)$ indicates that the curve has a λ parametrization.

2.4. Tangent Vector

The vector tangent to the curve was presented previously by Brown and Zingg [4]:

$$\dot{c}(s) = \xi \frac{\tau}{\|\tau\|}, \quad \tau = \begin{pmatrix} \mathbf{z} \\ -1 \end{pmatrix}, \quad \mathbf{z} = \left[\nabla_{\mathbf{q}} \mathcal{H}(\mathbf{q}, \lambda)\right]^{-1} \frac{\partial}{\partial \lambda} \mathcal{H}(\mathbf{q}, \lambda), \tag{8}$$

where $\mathbf{z} \in \mathbb{R}^N$, $\tau \in \mathbb{R}^{N+1}$, and $\xi = \pm 1$. According to Allgower and Georg [1], to ensure that the curve is traced in a consistent direction, ξ should be chosen such that the sign of

$$\det \begin{pmatrix} \nabla \mathcal{H}(\mathbf{q}, \lambda) \\ \dot{c}(s)^T \end{pmatrix}$$

is consistent for all $\lambda \in \Lambda$. However, if it is assumed that no bifurcation points are present, then we should choose $\xi = -1$ since this will give $\dot{\lambda}(s) < 0$.

An expression for the tangent vector under a λ parametrization can be developed by differentiating $\mathcal{H}(c(r)) = \mathbf{0}$ and using condition (7):

$$\nabla_{\mathbf{q}} \mathcal{H}(c(r)) \dot{\mathbf{q}}(r) = \frac{\partial}{\partial \lambda} \mathcal{H}(c(r)). \tag{9}$$

Note the very useful property

$$\dot{\mathbf{q}}(r) = \sqrt{\mathbf{z} \cdot \mathbf{z} + 1}\, \dot{\mathbf{q}}(s), \tag{10}$$

which allows for easy conversion between $\dot{\mathbf{q}}(s)$ and $\dot{\mathbf{q}}(r)$.

The tangent calculation presented in this section, if the algebra is performed exactly, is an exact expression for the tangent vector. However, when the calculation is performed in practice, the calculation can be inexact for several reasons:

1. The point at which the tangent is evaluated is typically an inexact solution to $\mathcal{H}(\mathbf{q}, \lambda) = \mathbf{0}$;
2. There may be some inaccuracy associated with the solution to the linear system; in the numerical studies presented in this paper, the linear system is solved inexactly using a Krylov solver, and the Jacobian-vector products are formed inexactly;
3. Round off error, though typically insignificant compared to the other two sources of error.

2.5. Curvature Vector Estimation

Direct methods have been developed to calculate the curvature vector $\ddot{c}(s)$ applicable to homotopies in low spatial dimensions [25, 30, 38, 41]. Though we have also developed a method for directly calculating curvature suitable for application to large sparse systems [5], this method is not used here as the work was unpublished at the time that the current manuscript was being prepared. However, many of the calculations presented in the current paper have been carried out using the direct method and can be found in the first author's PhD thesis [3].

Curvature vectors are estimated by applying finite-differencing to the tangent vector. The following centred-difference estimate is used when considering an arclength parametrization:

$$\ddot{c}\left(s_{i+\frac{1}{2}}\right) \approx \ddot{c}^*\left(s_{i+\frac{1}{2}}\right) = \frac{1}{\Delta s_i^*}\left(\dot{c}\left(s_{i+1}\right) - \dot{c}\left(s_i\right)\right), \tag{11}$$

$$s_{i+\frac{1}{2}} \approx \frac{1}{2}\left(s_{i+1} + s_i\right).$$

A similar expression is used when considering a λ parametrization.

Note that once the finite-difference step size $|\Delta\lambda|$ becomes sufficiently small, the error in the finite-difference approximation to $\ddot{c}(s)$ will begin to increase with decreasing step size. This can be explained by considering the error vectors $\mathbf{e}(s_i), \mathbf{e}(s_{i+1}) \in \mathbb{R}^N$ associated with the tangent vectors, which are independent of the estimated step size $\Delta s^* = s_{i+1} - s_i$. Then the curvature approximation can be written as

$$\ddot{c}^*(s) = \frac{1}{\Delta s^*}\left(\dot{c}\left(s_{i+1}\right) + \mathbf{e}\left(s_{i+1}\right) - \dot{c}\left(s_i\right) - \mathbf{e}\left(s_i\right)\right). \tag{12}$$

As Δs^* goes to 0, $\frac{1}{\Delta s^*}\left(\dot{c}\left(s_{i+1}\right) - \dot{c}\left(s_i\right)\right)$ approaches $\ddot{c}(s)$ but, since $\mathbf{e}(s)$ does not decrease with Δs, $\frac{1}{\Delta s^*}\left|\mathbf{e}\left(s_{i+1}\right) - \mathbf{e}\left(s_i\right)\right|$ will grow as Δs^* decreases and will eventually become larger than the truncation error. In effect, arbitrary accuracy cannot be achieved by taking Δs arbitrarily small, and the value of Δs corresponding to the maximum accuracy which can be achieved depends on how accurately the tangent is calculated. Assuming that the error associated with the tangent estimation can be modeled as random, propagation of this error into the curvature estimate will manifest as random noise.

2.6. Curvature

Curvature is a scalar metric calculated from the curvature vector. Several definitions are given for different types of curvature:

Definition 3. *Let $c : \mathbb{R} \to \mathbb{R}^M$, $(s) \mapsto c(s)$ be a C^2-differentiable curve in $\mathbb{R}^M$ with arclength parameter $s \in \mathbb{S}$, $\mathbb{S} \subset \mathbb{R}$. The total curvature [22] $\kappa : \mathbb{R} \to \mathbb{R}$, $(s) \mapsto \kappa(s)$ is defined as*

$$\kappa(s) = \sqrt{\ddot{c}(s) \cdot \ddot{c}(s)}, \tag{13}$$

and the partial curvature $\kappa_{\mathbf{q}} : \mathbb{R} \to \mathbb{R}$ is defined as

$$\kappa_{\mathbf{q}}(s) = \sqrt{\ddot{\mathbf{q}}(s) \cdot \ddot{\mathbf{q}}(s)}. \tag{14}$$

Definition 4. *Let $c : \mathbb{R} :\to \mathbb{R}^M \times \mathbb{R}$, $(r) \mapsto c(r)$ be a C^2-differentiable curve in $\mathbb{R}^M \times \mathbb{R}$ of the form $c(\mathbf{q}(r), \lambda(r))$ with parameter $r \in \mathbb{R}$ such that $\dot{\lambda}(r) = -1$. The total curvature with λ parametrization $\kappa_r : \mathbb{R} \to \mathbb{R}$, $(r) \mapsto \kappa(r)$ is defined as*

$$\kappa_r(r) = \sqrt{\ddot{\mathbf{q}}(r) \cdot \ddot{\mathbf{q}}(r)}. \tag{15}$$

2.7. Validation of the Curvature Estimation

The error in the finite-difference curvature estimations can be approximated numerically using Richardson extrapolation to estimate the converged curvature values, meaning the curvature estimated on an infinite-resolution grid, where the grid, in the present context, consists of the set of s or λ values, as appropriate, over which the curvature estimates are evaluated. To apply Richardson extrapolation, the convergence rate $p \in \mathbb{R}$ must first be calculated from the three finest converged functional values:

$$p = \frac{\ln\left(|f_3 - f_2| \,/\, |f_2 - f_1|\right)}{\ln(r)}, \tag{16}$$

where f_i is the functional of interest (curvature) at grid level i, where $i = 1$ is the finest grid and $i = 3$ the coarsest, and $r \in \mathbb{R}$ is the grid ratio. The grid ratio is equal to 2 for all studies in this paper. The convergence rate is then used to extrapolate the functional value on an infinite resolution mesh, denoted f_0, using the two finest grid levels:

$$f_0 = \frac{r^p f_1 - f_2}{r^p - 1}. \tag{17}$$

If the final three points are non-monotonic, or if the estimated convergence rate is less than 1, then the grid-converged functional is estimated from a first-order extrapolation by setting $p = 1$ in equation (17).

Validation is performed for each curvature study using grid sizes based on $|\Delta\lambda| = 0.02$, 0.04, and 0.08. For the $\kappa_{\mathbf{q}}$ calculation, this does not result in perfectly uniform grid refinement since the grid consists of the arclength values. The curvature estimates are performed every $|\Delta\lambda| = 0.02$ regardless of grid spacing. Due to the coarse grid spacing, Richardson extrapolation can only be performed for curvature values corresponding to λ values between 0.96 and 0.04. Since centered differencing is used, curvature values using the fine mesh correspond to λ values such as 0.95, 0.93, 0.91, etc, whereas curvature values estimated on the coarser grids correspond to λ values of 0.94, 0.92, etc. Since Richardson extrapolation must be applied to the curvature estimated at the same nodal locations on each grid level, the fine grid estimates are interpolated to $\lambda = 0.94$, 0.92, etc. using cubic splines. In the case of the $\kappa_{\mathbf{q}}$ estimates, $s(\lambda)$ is in addition interpolated to $\lambda = 0.95$, 0.93, etc. using cubic splines.

2.8. μ-Scaling

If a homotopy continuation algorithm is limited to local curve and tangent information, then the prediction of subsequent points along the curve is hindered if the curvature is significant. Step-length adaptation can be used to mitigate this effect locally by reducing the step size in regions of high curvature. However, step-length adaptation algorithms have limited effectiveness, especially if the curvature increases significantly over a short arc segment. The μ-scaling was developed as a means to improve the performance of step-length adaptation algorithms by distributing the curvature more evenly across the homotopy curve [4]. This scaling is equivalent to a global re-parametrization, which can be written as a coordinate transformation in λ:

$$\lambda \leftarrow \frac{\lambda\mu}{1 - \lambda + \mu\lambda}. \tag{18}$$

This transformation is equivalent to explicitly modifying the convex homotopy given by equation (3) to the following form:

$$\mathcal{H}(\mathbf{q}, \lambda_k) = (1 - \lambda_k)\mathcal{R}(\mathbf{q}) + \lambda_k \mu \mathcal{G}(\mathbf{q}) = \mathbf{0}, \tag{19}$$

$$k \in [0, m], \ \lambda_k \in \mathbb{R}, \ \lambda_0 = 1, \ \lambda_m = 0, \ \lambda_{k+1} < \lambda_k.$$

This modified homotopy equation is reminiscent of a technique known as the "γ-trick" [13, 27, 39], where a complex parameter is included in the equation for the purpose of avoiding potential bifurcation points. The difference is that our modification does not fundamentally alter the curve and exists solely for the purpose of re-distributing the curvature. As can be seen by the change of coordinates formula (18), $\mu > 1$ effectively compresses the domain near $\lambda = 0$ and stretches it near $\lambda = 1$, while $\mu < 1$ has the reverse effect.

In practice, it is convenient to apply μ-scaling as two components:

$$\mu = \mu_a \mu_u, \ \ \mu_a, \mu_u \in \mathbb{R}, \tag{20}$$

where μ_a is some fixed benchmark value that can be determined from a numerical algorithm or based on experience, and μ_u is user-supplied. In this study, the minimization algorithm of Brown and Zingg [4] is used to determine an appropriate value for μ_a. When using the dissipation operator with far-field boundary conditions (described in Section 6) as the homotopy system, μ_a is determined to be 0.7 for inviscid flow, 0.5 for laminar flow, and 0.1 for turbulent cases, each obtained from a single subsonic case over the ONERA M6 [37] wing.

3. Newton-Krylov Flow Solver

3.1. Computational Aerodynamics Flow Solver

With the exception of the first test case (Section 7.1), the numerical studies are performed using a Newton-Krylov-Schur parallel implicit flow solver based on a finite-difference [23] discretization applicable to multi-block structured grids. The discretization is based on the SBP-SAT [8, 9, 11, 21] approach, which uses Summation-By-Parts (SBP) operators to represent the discrete derivatives and Simultaneous Approximation Terms (SATs) to enforce the boundary conditions and couple the flow equations at block interfaces. The nonlinear numerical scheme is stabilized using the artificial dissipation operator of Pulliam [31], which was based on the earlier work of Jameson *et al.* [19].

The original inviscid flow solver flow of Hicken and Zingg [15] is used for all calculations in this paper though the flow solver has since been extended to viscous flows by Osusky and Zingg [29]. Turbulence is modeled using the Reynolds-averaged Navier-Stokes (RANS) equations with Spalart-Allmaras (SA) [40] turbulence model. The flow solver can be used for both subsonic and transonic operating conditions. For transonic cases, a first-order dissipation operator is included with a pressure sensor [15, 19] for shock capturing.

3.2. Newton's Method

Consider a nonlinear algebraic system of equations, represented by

$$\mathcal{F}(\mathbf{q}) = \mathbf{0}, \tag{21}$$

$$\mathcal{F} : \mathbb{R}^N \to \mathbb{R}^N, \ \mathbf{q} \in \mathbb{R}^N.$$

The update due to Newton's method, when applied to this system of equations, is calculated by solving the linear system of equations

$$\nabla \mathcal{F}^{(n)}(\mathbf{q}) \ \Delta \mathbf{q}^{(n)} = -\mathcal{F}\left(\mathbf{q}^{(n)}\right), \tag{22}$$

$$\Delta \mathbf{q}^{(n)} \equiv \mathbf{q}^{(n+1)} - \mathbf{q}^{(n)},$$

where $\nabla \mathcal{F}^{(n)} : \mathbb{R}^N \to \mathbb{R}^N$ is the Jacobian of $\mathcal{F}(\mathbf{q})$, defined as

$$\nabla \mathcal{F}_{i,j}(\mathbf{q}) \equiv \frac{\partial \mathcal{F}_i(\mathbf{q})}{\partial \mathbf{q}_j}, \tag{23}$$

which can be represented by a square matrix.

3.3. Solving the Distributed Linear System

To parallelize the flow solver, the domain is decomposed into blocks. Parallel preconditioning of the Krylov solver is performed using the Schur complement method [34] with block incomplete lower-upper (ILU) preconditioning applied to the domain blocks. The specific type of ILU factorization used in the current study is known as ILU(p) [34], where p is the fill level. The ILU(p) factorization is constructed based on an approximate Jacobian matrix using nearest neighbour nodes only. Since the Schur preconditioner can vary slightly throughout the Krylov solution process, a flexible variant of the Krylov solver GMRES is used, which is termed Flexible Generalized Minimal Residual, or FGMRES [34].

3.4. Jacobian-Free Newton-Krylov

Since the linear system (22) is solved to some relative tolerance $\tau_1^{(n)} \in \mathbb{R}$ using FGMRES, the actual Newton step is taken inexactly, and the update $\Delta \mathbf{q}$ does not satisfy equation (22) but does satisfy the inequality

$$\left\| \mathcal{F}\left(\mathbf{q}^{(n)}\right) + \nabla \mathcal{F}^{(n)}(\mathbf{q}) \ \Delta \mathbf{q}^{(n)} \right\| \leq \tau_1^{(n)} \left\| \mathcal{F}\left(\mathbf{q}^{(n)}\right) \right\|. \tag{24}$$

It is not always necessary to form and store the Jacobian matrix in a Newton-Krylov algorithm [20] since this matrix is not needed explicitly by a Krylov iterative method for solving linear systems, such as GMRES. What is needed by a Krylov solver are the Jacobian-vector products, which can be approximated using, for example, a finite-differencing approach [15, 28].

3.5. Solving the Nonlinear System

Newton's method has a high convergence rate but usually requires a good initial guess to be successful. A globally convergent method such as pseudo-transient continuation or homotopy continuation can be used to obtain a suitable starting guess for Newton's method. Hence the solution process is divided into two phases: the continuation phase and the inexact Newton phase.

4. Homotopy Design

4.1. Desired Properties of $\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)$

The suitability of a given $\mathcal{G}(\mathbf{q})$ as a homotopy system will depend on the curve-tracing methodology. An important consideration for curve-tracing efficiency is how accommodating the Jacobian of the homotopy is to the (preconditioned) linear solver. In the special case of the (flexible) GMRES-ILU(p) combination, it is important that the Jacobian matrix $\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)$ should be positive definite [34] and well-conditioned.

As discussed in Section 1, it is contrary to our objectives to construct homotopies which feature bifurcations. To ensure that no bifurcations exist, the determinant of $\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)$ should be maintained positive for all $\lambda \in \Lambda$, as explained in the following paragraphs.

Recall the condition, presented in Section 2.4, that the value of ξ used in the tangent calculation in equation (8) should be chosen such that the sign of

$$\det\begin{pmatrix} \nabla\mathcal{H}(\mathbf{q},\lambda) \\ \dot{c}(s)^T \end{pmatrix}$$

is consistent for all $\lambda \in \Lambda$. Considering equation (8), we can derive a relationship between this expression and $\det\left(\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)\right)$:

$$\begin{aligned}
\det\begin{pmatrix} \nabla\mathcal{H}(\mathbf{q},\lambda) \\ \dot{c}(s)^T \end{pmatrix} &= \det\begin{pmatrix} \nabla_{\mathbf{q}}\mathcal{H}(c(s)) & \frac{\partial}{\partial\lambda}\mathcal{H}(c(s)) \\ \xi\frac{1}{\|\tau\|}\mathbf{z}^T & -\xi\frac{1}{\|\tau\|} \end{pmatrix} \qquad (25)\\
&= \xi\frac{1}{\|\tau\|}\det\begin{pmatrix} \nabla_{\mathbf{q}}\mathcal{H}(c(s)) & \nabla_{\mathbf{q}}\mathcal{H}(c(s))\,\mathbf{z} \\ \mathbf{z}^T & -1 \end{pmatrix} \\
&= \xi\frac{1}{\|\tau\|}\det\begin{pmatrix} \nabla_{\mathbf{q}}\mathcal{H}(c(s)) & 0 \\ \mathbf{z}^T & \mathbf{z}\cdot\mathbf{z}+1 \end{pmatrix}\det\begin{pmatrix} \mathcal{I} & \mathbf{z} \\ \mathbf{0}^T & -1 \end{pmatrix} \\
&= -\xi\frac{1}{\|\tau\|}(1+\mathbf{z}\cdot\mathbf{z})\det\left(\nabla_{\mathbf{q}}\mathcal{H}(c(s))\right), \qquad (26)
\end{aligned}$$

We observe that a change of sign in the determinant of $\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)$ will cause ξ and hence, considering equation (8), $\dot{\lambda}(s)$ to change sign, indicating the presence of a Hopf bifurcation somewhere along the curve segment.

Assuming that we are only interested in studying dynamically stable flow solutions, all eigenvalues of the converged flow solution may be assumed to have positive (by convention) real part. Since the determinant is equal to the product of the eigenvalues, the determinant of $\nabla\mathcal{R}(\mathbf{q})$ is positive at the converged flow solution. A desirable property of $\mathcal{G}(\mathbf{q})$ is therefore that it should induce $\det\left(\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)\right) > 0$ for all $\lambda \in \Lambda$, which, since the determinant is equal to the product of the eigenvalues, can be assured if all real eigenvalues of $\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q},\lambda)$ are strictly positive for all λ. (Since complex eigenvalues always occur in complex conjugate pairs for real-valued matrices, complex eigenvalues do not affect the sign of the determinant.)

4.2. Curve Traceability

The implicitly-defined homotopy curve is nonlinear and can be traced with limited accuracy with a given numerical continuation method. The homotopy should be designed to facilitate curve tracing using the continuation algorithm. This is discussed further in Section 5.2.

4.3. Convex Homotopy System Design Objectives

In the context of the Newton-Krylov framework, the suitability of a homotopy system $\mathcal{G}$ for use in a homotopy continuation algorithm depends on how well each of the following criteria are met:

1. The matrix $\nabla\mathcal{G}(\mathbf{q})$ improves the performance of the linear solver when added to $\nabla\mathcal{R}(\mathbf{q})$
   - Most of the computational work is within the linear solver, so the efficiency of the linear solver is crucial to the efficiency of the algorithm
2. All real-valued eigenvalues of $\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{q}, \lambda)$ are strictly positive for all λ
   - As per Section 4.1
3. The solution to $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ is known or easily obtainable
   - The algorithm would be redundant if $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ is as difficult to solve as $\mathcal{R}(\mathbf{q}) = \mathbf{0}$
4. The solution to $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ is unique
   - Otherwise, we risk introducing bifurcations
5. The homotopy connecting $\mathcal{G}^{-1}(\mathbf{0})$ and $\mathcal{R}^{-1}(\mathbf{0})$ is easy to trace.
   - Though difficult to predict, this directly impacts the efficiency, robustness, and even the viability of the curve-tracing algorithm

5. Methods for Numerical Analysis

5.1. Surrogate Curves

Since the curves representing the homotopies exist in higher dimensional real space and cannot be visualized, one-dimensional surrogate curves can be used to assess the performance of the homotopy continuation algorithms. For example, the values of the lift coefficients (C_l for two-dimensional flow, C_L for three-dimensional flow) and drag coefficients (C_d for two-dimensional flow, C_D for three-dimensional flow) calculated along the curve can be used as one-dimensional surrogates for the curve. These surrogates are not expected to give a realistic impression of the features of the curve such as curvature but can be used to roughly evaluate the effectiveness of various curve tracing or curve prediction tools or to assess how similar two different homotopies might be.

5.2. Curvature

While it is clear that curvature is an important metric for evaluating curve traceability, the apparent curvature, and hence traceability, also depends on the parametrization. For example, if it is assumed that traversing is performed with a constant step-length, where the step-length is measured with respect to an arclength parametrization, then it is important to consider $\kappa_{\mathbf{q}}$. However, if it is assumed that the curve is traced with constant $\Delta\lambda$ then it is more relevant to consider κ_r. Since the use of λ as the parameter controlling the deformation is a matter of convenience and not performance, curve tracing algorithms are generally designed to attempt to maintain a relatively consistent Δs. However, in our experience, it is usually advantageous to enforce upper and lower bounds on $\Delta\lambda$, as well as upper and lower bounds on the factor that $\Delta\lambda$ can change by between successive iterations, so some elements of a λ parametrization can be present in practice. Hence, both parametrizations are typically relevant.

Since $\kappa_{\mathbf{q}}$ is (mostly) independent of the parametrization, this is a useful tool for assessing

the suitability of a homotopy system for use in a continuation algorithm under the assumption that the curve can be re-parametrized. However, it is not a good metric for directly comparing two homotopies because it assumes both curves are being traced with the same step size Δs and does not take into account that, under this condition, more steps would be needed if the curve is "longer". To establish a more appropriate metric, consider the Taylor expansion around some s_0:

$$\mathbf{q}(s_0 + \Delta s) = \mathbf{q}(s_0) + \Delta s \dot{\mathbf{q}}(s_0) + \frac{1}{2}\Delta s^2 \ddot{\mathbf{q}}(s_0) + O\left(\Delta s^3\right). \tag{27}$$

If a predictor is formed using only $c(s)$ and $\dot{c}(s)$, then, neglecting $O\left(\Delta s^3\right)$ terms, the norm of the error $e \in \mathbb{R}$ resulting from the curvature is

$$e = \sqrt{\left(\frac{1}{2}\Delta s^2\right)^2 \ddot{\mathbf{q}}(s) \cdot \ddot{\mathbf{q}}(s)} = \frac{1}{4}\Delta s^2 \kappa_{\mathbf{q}}. \tag{28}$$

If it is assumed that the curve is always traced with the same number n_s of equally spaced (in Δs) steps then $s_{\text{tot}} = n_s \Delta s$ and hence equation (28) becomes

$$e = \frac{1}{4n_s^2} s_{\text{tot}}^2 \kappa_{\mathbf{q}}. \tag{29}$$

The actual value of n_s is irrelevant for comparison and so $s_{\text{tot}}^2 \kappa_{\mathbf{q}}$ is the traceability metric considered when assuming an arclength parametrization. This quantity can be plotted against λ to quantify how traceability varies with λ. To assess how traceability varies (nearly) independently of the parametrization, the traceability metric could instead be plotted against s/s_{tot}. The values of Δs and s_{tot} can be estimated numerically as

$$\Delta s_i \approx \sqrt{\|\mathbf{q}_i - \mathbf{q}_{i-1}\|^2 + |\lambda_i - \lambda_{i-1}|^2}, \tag{30}$$

$$s_{\text{tot}} \approx \sum_{i=1}^{n_s} \Delta s_i. \tag{31}$$

While the traceability metrics presented in this section are useful for comparing homotopies on a given mesh under certain flow conditions, the curvature can scale in complicated ways with mesh size, local grid refinement, and the state variables $\mathbf{q}$. Hence the traceability metrics should not, in general, be used to compare traceability across meshes or under different flow conditions. An exception to this is if the mesh is refined in a consistent way; such a study is performed in Section 7.5.

5.3. Linear System Convergence

The amount of curvature gives an indication of how many iterations will be required by the continuation algorithm. However, the cost of each iteration can vary significantly during the continuation process. For this reason, it is important to develop a profile for the convergence behaviour of the linear solver as a function of the continuation parameter. In this case, the linear solver is Schur-preconditioned FGMRES with block ILU(p).

In addition to curvature profiles, we would like to profile the FGMRES convergence behaviour using an accessible metric. A simple way to profile FGMRES convergence is simply to plot the number of linear iterations taken to converge the linear system residual below a specified relative tolerance τ_l. However, this assumes that the ideal τ_l is known and it is preferable to develop a metric which is independent of the τ_l value.

The linear solver convergence behaviour when solving a linear system of the form $\mathcal{A}\mathbf{x} = \mathbf{b}$ with FGMRES depends on both $\mathcal{A}$ and $\mathbf{b}$ and is an area of ongoing research [42]. Under the assumption that $\mathcal{A}$ is diagonalizable, an upper bound on the convergence profile has been known since the original paper on GMRES by Saad and Schultz [35]. For an application of GMRES that has reached the kth iteration, this convergence bound is given by the rather intractable expression

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \leq \kappa_n(\mathcal{V}) \min_{p\in\mathbb{P}_k, p(0)=1} \max_{\beta\in\mathcal{E}(\mathcal{A})} |p(\beta)|, \tag{32}$$

where $\mathbf{x}_k$ is the solution estimate at the kth iterate, and $\mathbf{r}_k \equiv \mathcal{A}\mathbf{x}_k - \mathbf{b}$ is the corresponding linear system residual. Additionally, $\kappa_n(\mathcal{V}) \equiv \|\mathcal{V}\|_2 \left\|\mathcal{V}^{-1}\right\|_2$ is the condition number of $\mathcal{V}$, $\mathcal{V}$ is the eigenvector matrix of $\mathcal{A}$, $\mathcal{E}$ is the set of eigenvalues of $\mathcal{A}$, and $\mathbb{P}_k$ is the space of kth degree polynomials. However, this is not necessarily a sharp convergence bound, as investigated by authors such as Embree [10]. In fact, Greenbaum *et al.* [12] have shown that for any non-increasing sequence, it is possible to construct $\mathcal{A}$ and $\mathbf{b}$ such that the sequence formed by the residual $\|\mathbf{r}_k\|$ matches this sequence.

Since a reliable method for predicting and profiling convergence of our preconditioned FGMRES linear solver does not appear to currently exist, we rely on a posteriori numerical analysis. We attempt to apply the model

$$R_k \equiv \frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \approx C\gamma^k, \tag{33}$$

with $C, \gamma \in \mathbb{R}$. Setting $C = 1$ satisfied the initial condition $R_k = 1$. However, since GMRES convergence often does not exhibit its asymptotic behaviour in the initial iterations [10], it may be preferable to consider both C and γ as variables, which can be solved for simultaneously by applying least-squares regression to the sequence

$$\ln R_k = \ln C + k \ln \gamma, \tag{34}$$

ignoring data from the early iterations. This may give a value of γ which is more representative of the asymptotic convergence rate. However, when solving the linear system to a fairly relaxed tolerance such as $\tau_l = 0.01$, the asymptotic region may not be prevalent, and so accurately modeling the asymptotic region of convergence may be less important. In the present study, $C = 1$ is assumed in the model and all R_k data is used in the regression. The effectiveness of this model for assessing linear solver performance is investigated numerically in Section 7.2, including the coefficient of determination R^2 and qualitative comparison of the model with the convergence data.

6. Some Specific Homotopy Systems

The operators presented in this section have been designed for use as homotopy systems with special consideration for the computational fluid dynamics flow solver of Hicken and Zingg [15]

and Osusky and Zingg [29]. It is expected that similar operators will also be effective for application to other equation solvers where the nonlinear system of equations has similar properties. In particular, these operators could be adapted to homotopy continuation algorithms for other computational fluid dynamics solvers.

The operators are designed such that the homotopies satisfy the first four requirements listed in Section 4.3. That is, they result in good performance in the linear solver for all $\lambda \in \Lambda$, the eigenvalues of the Jacobian of $\mathcal{G}(\mathbf{q})$ are positive, $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ can be solved easily, and $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ has a unique solution. Existence of the curve is verified numerically, and the curvature profiles are investigated numerically.

6.1. The Diagonal Operator

A linear operator $\mathcal{G}(\mathbf{q})$ with the property $(\mathcal{G}(\mathbf{q}))_{[i]} = g_i \mathbf{q}_{[i]}$, $g_i > 0$, $g_i \in \mathbb{R}$ is a suitable homotopy system because the Jacobian of this system is a diagonal positive-definite matrix. As such, the Jacobian is nonsingular, can be inverted trivially, and improves the diagonal dominance of the Jacobian when added to the discrete flow equations. Because the Jacobian is diagonal, this homotopy system will be referred to as the "Diagonal" operator.

By analogy to the Jacobian formed when applying a pseudo-transient method, the equation blocks of this homotopy system for the three-dimensional RANS equations take the form:

$$\mathcal{G}_{(i)}(\mathbf{q}) = \left(\frac{1 + \mathcal{J}_{[i]}^{\frac{1}{D}}}{\mathcal{J}_{[i]}}, \frac{1 + \mathcal{J}_{[i]}^{\frac{1}{D}}}{\mathcal{J}_{[i]}}, \frac{1 + \mathcal{J}_{[i]}^{\frac{1}{D}}}{\mathcal{J}_{[i]}}, \frac{1 + \mathcal{J}_{[i]}^{\frac{1}{D}}}{\mathcal{J}_{[i]}}, 1 + \mathcal{J}_{[i]}^{\frac{1}{D}} \right) \mathbf{q}_{(i)}, \tag{35}$$

where $\mathcal{J}$ is the metric Jacobian resulting from transforming the coordinates of the block-structured mesh from physical space to computational space and can be interpreted as the reciprocal of a grid cell volume (see, for example, Pulliam and Zingg [32]). The rounded brackets in the subscript (i) indicate the sub-vector corresponding to the ith grid node, and the square brackets $[i]$ indicate the ith component of the vector. The parameter D is the number of spatial dimensions and is equal to 3 for three-dimensional flows. For two-dimensional flows, D is set to 2, and the equation block size is reduced by one by deleting the fourth entry of equation (35), since this entry corresponds to the conservation of momentum equation in the third spatial direction. This homotopy system is similar in construction to the pseudo time operator used in pseudo-transient continuation algorithms. It varies from the operator found in the pseudo-transient continuation algorithm of Hicken and Zingg [15] only in that it does not contain a factor of the reference time step Δt_{ref}.

While this operator has an easily invertible Jacobian, the solution to $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ is $\mathbf{q} = \mathbf{0}$, which includes zero density and is non-physical. This is obviously a poor choice of homotopy system and so it is preferable to apply this operator as a warm-started homotopy operator as described in Section 6.3. This formulation allows for any specified value of $\mathbf{q}$, such as far-field conditions, to satisfy the modified homotopy system without affecting the system Jacobian. This is the approach taken in this paper, with a vector of far-field conditions used to construct the source term vector.

6.2. The Dissipation Operator

Scalar artificial dissipation operators such as presented by Jameson *et al.* [19] are often added to the discrete flow equations for numerical stability, and this dissipation scheme is used for the

studies in this paper. Adding such operators to the discrete flow equations can have a highly stabilizing effect and can improve the performance of the preconditioned linear solver by increasing the diagonal dominance of the Jacobian matrix. Hence, dissipation operators are suitable for application as a homotopy system. This has been previously acknowledged by Hicken *et al.* [14], who investigated the efficiency of a simple homotopy continuation algorithm using as the homotopy system the second-difference dissipation operator of Jameson *et al.* [19] absent the pressure sensor used for shock capturing.

The second-difference dissipation operator is given explicitly by

$$\mathcal{D}^{(2)}(\mathbf{q}) = \Delta^T C \Delta \mathbf{q}, \tag{36}$$

$$\Delta = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & -1 & 1 & & \\ & & & \ddots & \ddots & \\ & & & & -1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} d_{[\frac{1}{2}]} & & & & \\ & d_{[\frac{3}{2}]} & & & \\ & & \ddots & & \\ & & & d_{[N-\frac{3}{2}]} \end{bmatrix},$$

$$\Delta \in \mathbb{R}^{(N-1)\times N}, \quad C \in \mathbb{R}^{(N-1)\times(N-1)},$$

$$d_{[i+\frac{1}{2}]} = \frac{1}{2}\left(d_{[i]} + d_{[i+1]}\right), \quad d_{[i]} = \left|u_{[i]}\right| + a_{[i]}.$$

where a is the sound speed, u the fluid velocity. Note the u and a can both be written in terms of $\mathbf{q}$, making the operator nonlinear. For clarity of presentation, the one-dimensional scalar (eg. assuming one equation, and hence one variable, per grid point) version of the operator is presented. The operator can be applied to three-dimensional vector-valued equations by applying the operator given by equation (36) to each equation and in each spatial direction. Details are given by Pulliam and Zingg [32]. In addition, for three-dimensional flows, the equation for d is replaced by

$$d_{[i+\frac{1}{2}]} = \frac{1}{2}\left(d_{[i]} + d_{[i+1]}\right), \quad d_{[i]} = \left[\mathcal{J}^{-1}\left(|U_n| + a\sqrt{\xi_x^2 + \eta_y^2 + \zeta_z^2}\right)\right]_{[i]}, \tag{37}$$

where U_n is the contra-variant velocity and ξ_x, η_y, and ζ_z are related to the coordinate transformation [32].

That Δ defines a mapping from $\mathbb{R}^N$ to $\mathbb{R}^{N-1}$ is sufficient to identify that the rank of the $\mathcal{D}^{(2)}$ is at most $N-1$. Therefore $\mathcal{D}^{(2)}$ is singular and, as discussed previously, it is desired that the homotopy system should be nonsingular to ensure that the curve representing the homotopy deformation is unique. This can be remedied by augmenting $\mathcal{D}^{(2)}$ with pseudo boundary conditions.

Pseudo boundary conditions for $\mathcal{D}^{(2)}$ are formulated using the SAT approach to be consistent with the SATs used for the flow equations. The scalar one-dimensional version of the operator is analyzed for clarity of presentation. By analogy to the diffusion equation, the following form of the equation, including boundary conditions, is assumed:

$$\mathcal{G}(\mathbf{q}) = \mathcal{D}^{(2)}\mathbf{q} + \mathcal{B}(\mathbf{q}), \tag{38}$$

$$\mathcal{B}(\mathbf{q}) = \mathrm{diag}\left(\sigma_{\mathrm{L}}\left(\mathbf{q}_{[1]} - q_{\mathrm{L}}\right), 0, \ldots, 0, \sigma_{\mathrm{R}}\left(\mathbf{q}_{[N]} - q_{\mathrm{R}}\right)\right),$$

$$\mathcal{B} : \mathbb{R}^N \to \mathbb{R}^N, \quad \sigma_{\mathrm{L}}, \sigma_{\mathrm{R}}, q_{\mathrm{L}}, q_{\mathrm{R}} \in \mathbb{R}.$$

At a domain boundary, q_L and q_R are boundary conditions. At grid block interfaces, they are the flow values at the same point in physical space but corresponding to the adjacent block.

Since $\mathcal{D}^{(s)}$ is not an SBP operator, it is not straightforward to derive the stability condition on the SATs using the usual energy method. However, ignoring the dependence of d on $\mathbf{q}$ for simplicity, it is possible to determine some necessary conditions on the SATs by analysis of the pseudo-linear operator representing the Jacobian. For analysis, this operator is most conveniently expressed by giving the expression for the rows in the one-dimensional case:[1]

$$\mathcal{G}^{(l)}_{[i,:]} = \begin{cases} \left(\sigma_L + d_{[\frac{3}{2}]}, -d_{[\frac{3}{2}]}, 0, 0, \ldots, 0\right) & i = 1, \\ \mathrm{trid}_i\left(-d_{[i-\frac{1}{2}]}, d_{[i-\frac{1}{2}]} + d_{[i+\frac{1}{2}]}, -d_{[i+\frac{1}{2}]}\right) & i = 2, \ldots, N-1, \\ \left(0, \ldots, 0, 0, -d_{[N-\frac{1}{2}]}, d_{[N-\frac{1}{2}]} + \sigma_R\right) & i = N. \end{cases} \tag{39}$$

Consider the subtraction of the sum of the absolute values of the off-diagonal elements of the matrix from the absolute value of the diagonal elements:

$$\left|\mathcal{G}^{(l)}_{[i,i]}\right| - \sum_{j \neq i} \left|\mathcal{G}^{(l)}_{[i,j]}\right| = \begin{cases} \sigma_L & i = 1, \\ 0 & i = 2, \ldots, N-1, \\ \sigma_R & i = N. \end{cases} \tag{40}$$

Since $\mathcal{G}^{(l)}$ is easily verified to be irreducible, the necessary and sufficient condition for $\mathcal{G}^{(l)}$ to be irreducibly row-diagonally dominant is

$$\sigma_R \geq 0, \ \sigma_L \geq 0, \ \sigma_L + \sigma_R > 0. \tag{41}$$

Since irreducibly row-diagonally dominant systems are nonsingular [34], condition (41) is also a sufficient condition for non-singularity of $\mathcal{G}^{(l)}$.

By symmetry, and comparison with the well-known diffusion operator, it is inferred that the conditions for well-posedness are given by $\sigma_L = d_1$ and $\sigma_R = d_N$ for the scalar one-dimensional version of the operator. The extension to three-dimensional vector-valued systems is accomplished by similarly constructing the SATs in each direction and for each equation.

The boundary conditions q_L and q_R do not affect the conditioning of the linear system and can be chosen based on other criteria. One benefit to setting q_L and q_R to far-field conditions at all domain boundaries (including boundaries which are interpreted as solid surfaces for the physical problem) is that $\mathcal{G}(\mathbf{q}_{ff}) = \mathbf{0}$, where $\mathbf{q}_{ff} \in \mathbb{R}^N$ is the vector consisting of the far-field values in all corresponding elements. This ensures that the third criterion of Section 4.3 is satisfied as $\mathcal{G}(\mathbf{q}) = \mathbf{0}$ has a known solution. Since these boundary conditions are treated as part of the homotopy system, the actual boundary conditions of $\mathcal{H}(\mathbf{q}, \lambda)$ can be interpreted as gradually progressing from far-field everywhere at $\lambda = 1$ to the "correct" physical boundary conditions at $\lambda = 0$ with a non-physical combination of the two boundary condition types when $1 > \lambda > 0$.

Use of "flow-imitative" boundary conditions is also considered, where the boundary conditions are set at the solid surfaces, far-field boundaries, and symmetry planes which imitate the SATs corresponding to the flux terms of $\mathcal{R}(\mathbf{q})$. The motivation for including such boundary conditions is to attempt to produce a homotopy system which has solution closer to the solution of $\mathcal{R}(\mathbf{q}) = \mathbf{0}$, potentially leading to a shorter or somehow "easier" homotopy path. Though the solution is not immediately known for this choice of boundary conditions, from our experience it can be obtained with minor computational effort from a few inexact Newton iterations.

[1] In this context, $\mathrm{trid}_i(x, y, z)$ refers to the ith row of a matrix with x at the $i-1$st entry, y at the ith entry, z at the $i+1$st entry, and zeros everywhere else.

6.3. Warm-Started Homotopy Systems

The concept of "warm-starting" a nonlinear algorithm applies to fixed-point methods such as Newton's method or the pseudo-transient continuation method where a good initial guess $\mathbf{q}_0$ can reduce the total number of iterations taken by the algorithm and thus reduce the total CPU time needed for convergence. Global homotopy continuation can naturally take an initial guess $\mathbf{q}_0$, but convex homotopies cannot because $\mathbf{q}_0$ will not correspond to a point on the curve. However, an effect similar to warm-starting can be obtained in the context of convex homotopy for a given homotopy system $\mathcal{G}$ by constructing a modified homotopy system

$$\mathcal{G}^* (\mathbf{q}) = \mathcal{G} (\mathbf{q}) - \mathcal{G} (\mathbf{q}_0) \tag{42}$$

since $\mathbf{q} = \mathbf{q}_0$ is a solution to $\mathcal{G}^* (\mathbf{q}) = \mathbf{0}$.

7. Some Numerical Studies of Homotopies

The numerical studies are carried out in the context of the flow solver described in Section 3. The homotopy systems investigated are the convex homotopy with the dissipation operator using far-field boundary conditions, the dissipation operator using flow-imitative boundary conditions, and the diagonal operator with a source term based on far-field conditions, as well as global homotopy. These are indicated as "Diss - ff", "Diss - flow", "Diag", and "Global" respectively on the plots. Richardson extrapolation is abbreviated as RE.

7.1. Surrogate Curves for some One-Dimensional Inviscid Homotopies

Two cases are considered, one of which is subsonic and the other transonic. Both cases consist of airflow through a converging-diverging nozzle under conditions such that compressibility cannot be neglected. The shape of the nozzle $S(x)$ for both cases is given by

$$S(x) = \begin{cases} 1 + 1.5\left(1 - \frac{x}{5}\right)^2 & 0 \le x \le 5 \\ 1 + 0.5\left(1 - \frac{x}{5}\right)^2 & 5 \le x \le 10. \end{cases} \tag{43}$$

For both cases, the air is considered to be a perfect gas with ideal gas constant $R = 287\text{N} \cdot \text{m} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$, heat capacity ratio $\gamma = 1.4$, total temperature $T_0 = 300\text{K}$, and total inlet pressure $p_{01} = 100\text{kPa}$. The critical area is $S^* = 0.8$ for the subsonic case and $S^* = 1$ for the transonic case. The critical area is used to calculate the Mach number at the inlet in this case and can also be used to calculate the Mach number for all x. Both problems are described by Pulliam and Zingg [32], who in turn reference Hirsch [18]. More details of the problem can be found in either textbook, including the relationship between S^* and the Mach number.

To discretize the governing system of partial differential equations we have used the SBP-SAT approach with 200 equally spaced nodes and an interface at $x = 6\text{m}$, where coupling across the interface is also enforced using the SAT approach. The location of the interface has been chosen more or less arbitrarily near the mid-point of the domain. The implementation is in Matlab using an LU decomposition to solve the linear system. Flow cases for both the subsonic and transonic conditions of Pulliam and Zingg [32] are considered.

The homotopies considered are the convex homotopy with the dissipation operator, where the pseudo boundary conditions match the physical inlet and outlet boundary conditions of the flow, and the global homotopy, where $\mathbf{q}_0$ is set to the physical boundary conditions at the domain

boundaries and interpolated based on these values at points interior to the domain. The surrogate for the deformation in this case is the pressure profile over the spatial domain $x \in [0, 10]$, $x \in \mathbb{R}$. It can be seen from Figure 1 that for the subsonic case the global homotopy curve is much more traceable than the convex homotopy curve with the dissipation operator. However, the convex homotopy curve is far more traceable in the transonic case. When applying the global homotopy to the transonic case, the predictor-corrector method was found to diverge from the curve unless a very small step size (approximately $|\Delta\lambda| = 0.001$) was taken, even for this relatively simple problem.

The performance of the linear solver can vary greatly from the continuation phase to the inexact Newton phase depending on the continuation algorithm. Previously, based on an implementation of the one-dimensional linear convection equation, Hicken *et al.* [14] investigated the evolution of the condition number and eigenvalue spectrum with respect to the continuation parameter for a dissipation-based continuation strategy and compared the results to a pseudo-transient continuation algorithm. A similar study is performed here for the subsonic converging-diverging nozzle problem, and a similar trend is observed. Evolution of the condition number κ_n is shown for the respective Jacobian matrices of the convex homotopy with dissipation operator and the pseudo-transient continuation method in Figure 2. For pseudo-transient continuation, the condition number is plotted against the Courant number C_n, which is used to define the spatially-varying time step Δt using the formula

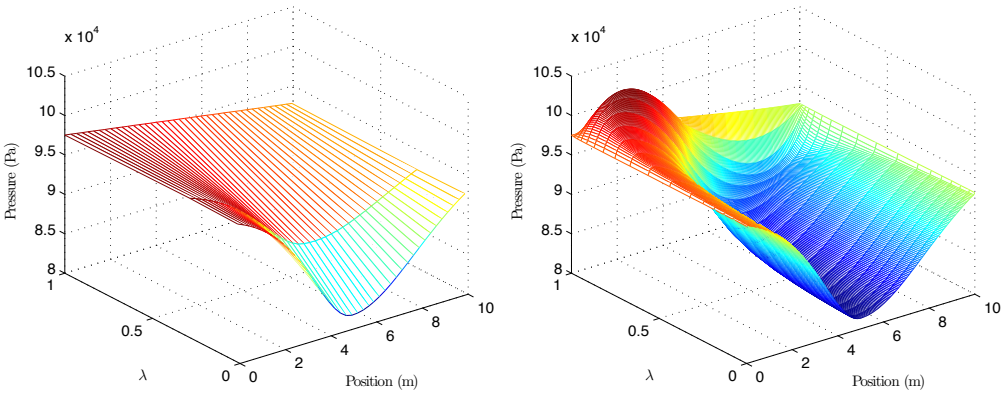
$$\Delta t = \frac{C_n \Delta x}{(|u| + a)}, \tag{44}$$

where u is the local fluid velocity and a the local speed of sound.

It is apparent from Figure 2 that reducing the Courant number (and hence the time step) has the effect of improving the condition number, whereas adding the dissipation operator to the system actually worsens the conditioning. To study these effects, the eigenvalue spectra are investigated for different λ and C_n in Figures 3 through 5. For pseudo-transient continuation, when the Courant number is decreased, the eigenvalues migrate away from the imaginary axis so the difference in magnitude of the smallest and largest in magnitude eigenvalues decreases, resulting in improved conditioning. For the convex homotopy with dissipation operator, increasing λ compresses the eigenvalue spectrum to and stretches it along the real axis. Though pseudo-boundary conditions have been applied to the dissipation operator, the smallest in magnitude eigenvalue at $\lambda = 1$ is near to zero, resulting in a poorly-conditioned system. However, what is important for the convergence of Krylov iterative solvers is not the conditioning of the Jacobian matrix itself but the conditioning of the preconditioned matrix. When increasing λ, the diagonal dominance and definiteness of the linear system is gradually improved, which can significantly improve the quality of the ILU decomposition when using the ILU(p) algorithm [34]. As a result, the conditioning of the preconditioned linear system can still be quite good. Since the one-dimensional problem studied here results in a narrow banded matrix for which a complete LU factorization can easily be formed, this is not a suitable problem for studying linear system performance and so we leave these studies for the two- and three-dimensional cases.

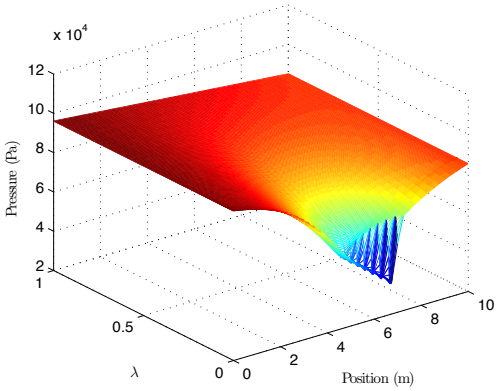
7.2. Curvature Profiles for a Two-Dimensional Inviscid Subsonic Flow

The test case for this study is an inviscid subsonic flow over the NACA 0012 airfoil. The grid consists of 18 blocks divided evenly into 15390 nodes. The operating conditions are Mach 0.3 and an angle of attack of 1°. Surrogate curves and curvature profiles are generated for the

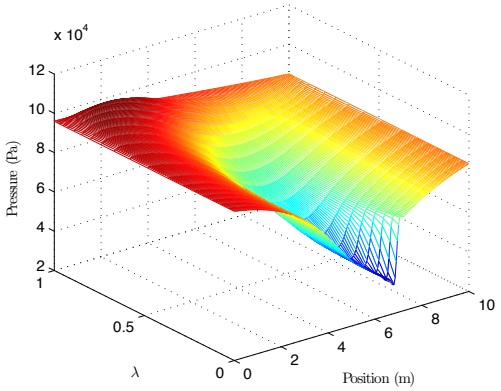


(a) Global homotopy, subsonic flow

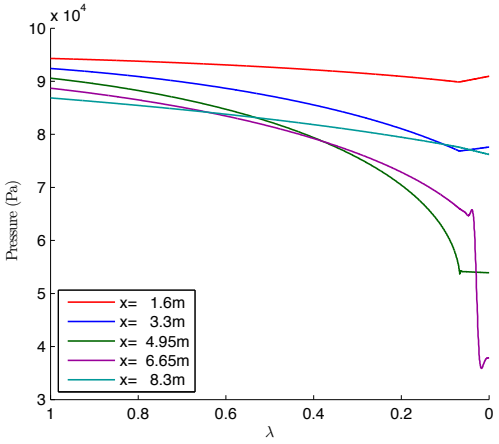
(b) Convex homotopy, subsonic flow



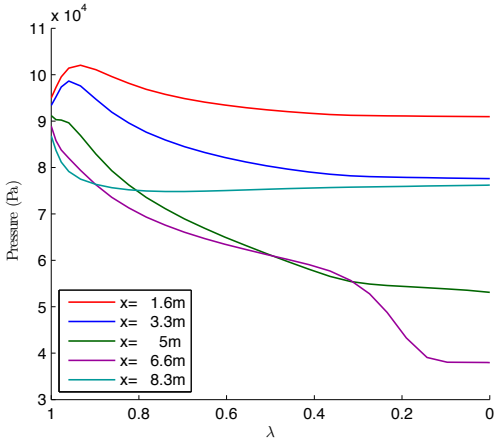
(c) Global homotopy, transonic flow



(d) Convex homotopy, transonic flow

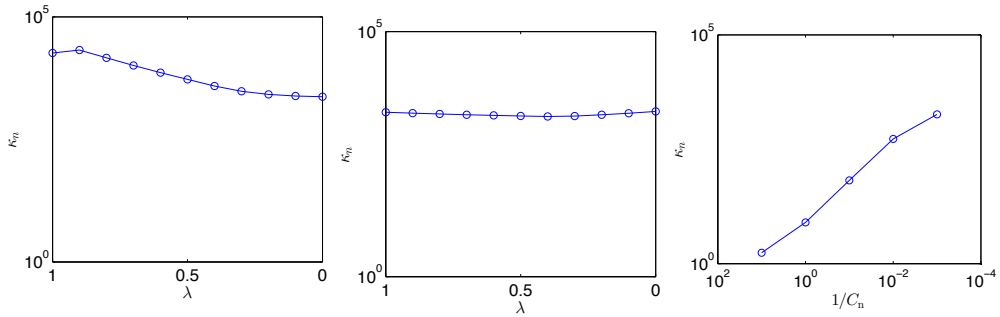


(e) Global homotopy, transonic flow



(f) Convex homotopy, transonic flow

Figure 1: Surrogate curves for the global and convex homotopies with the dissipation homotopy system for a one-dimensional converging-diverging nozzle problem under subsonic and transonic conditions

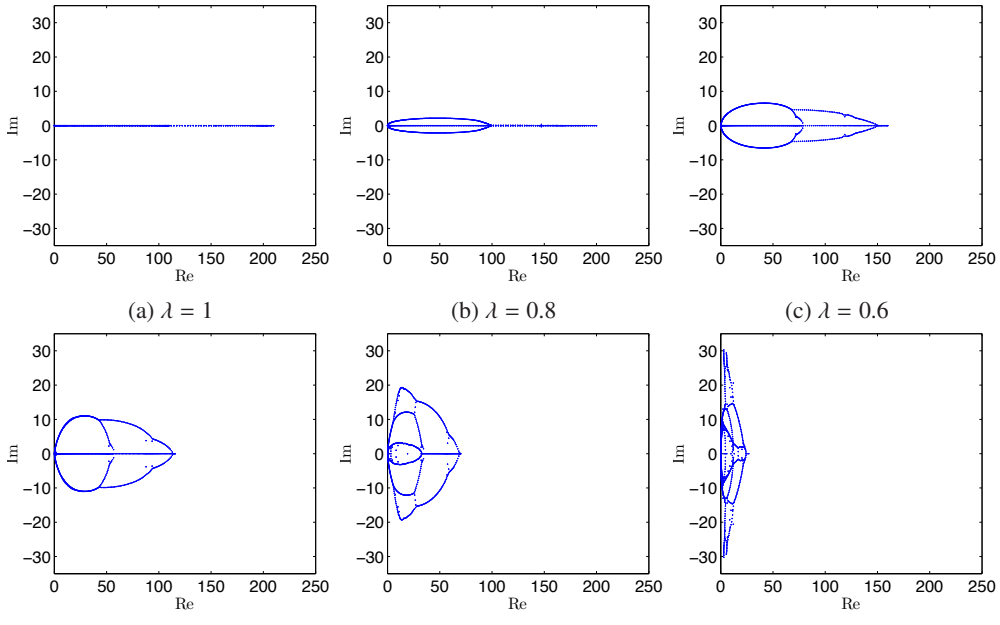


(a) Convex homotopy with dissipation operator

(b) Global homotopy

(c) Pseudo-transient continuation

Figure 2: Condition number evolution for three continuation methods for the subsonic converging-diverging nozzle



(a) $\lambda = 1$

(b) $\lambda = 0.8$

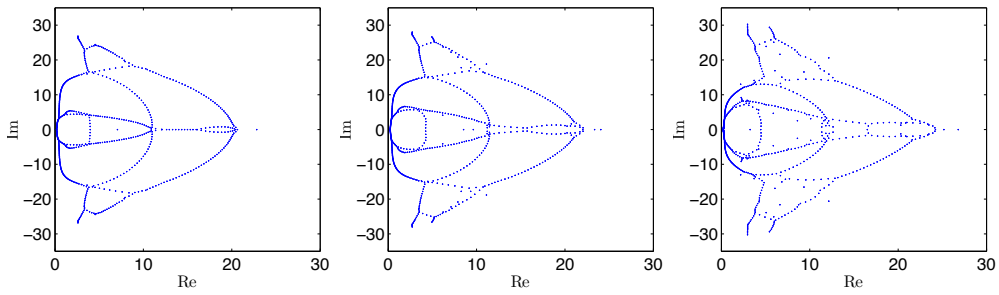
(c) $\lambda = 0.6$

(d) $\lambda = 0.4$

(e) $\lambda = 0.2$

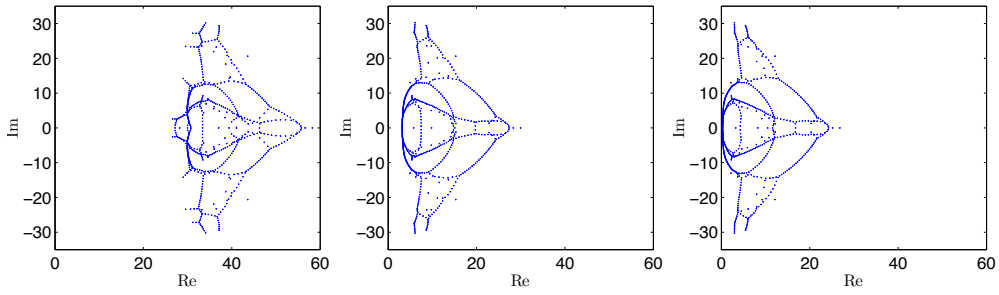
(f) $\lambda = 0$

Figure 3: Evolution of the eigenvalue spectrum of the Jacobian matrix for the convex homotopy with dissipation operator for the subsonic converging-diverging nozzle case



(a) $\lambda = 1$ (b) $\lambda = 0.5$ (c) $\lambda = 0$

Figure 4: Evolution of the eigenvalue spectrum of the Jacobian matrix for the global homotopy for the subsonic converging-diverging nozzle case



(a) $1/C_{\mathrm{n}} = 1$ (b) $1/C_{\mathrm{n}} = 0.1$ (c) $1/C_{\mathrm{n}} = 0$ (Newton)

Figure 5: Eigenvalue spectrum of the Jacobian matrix for pseudo-transient continuation with different Courant numbers for the subsonic converging-diverging nozzle case

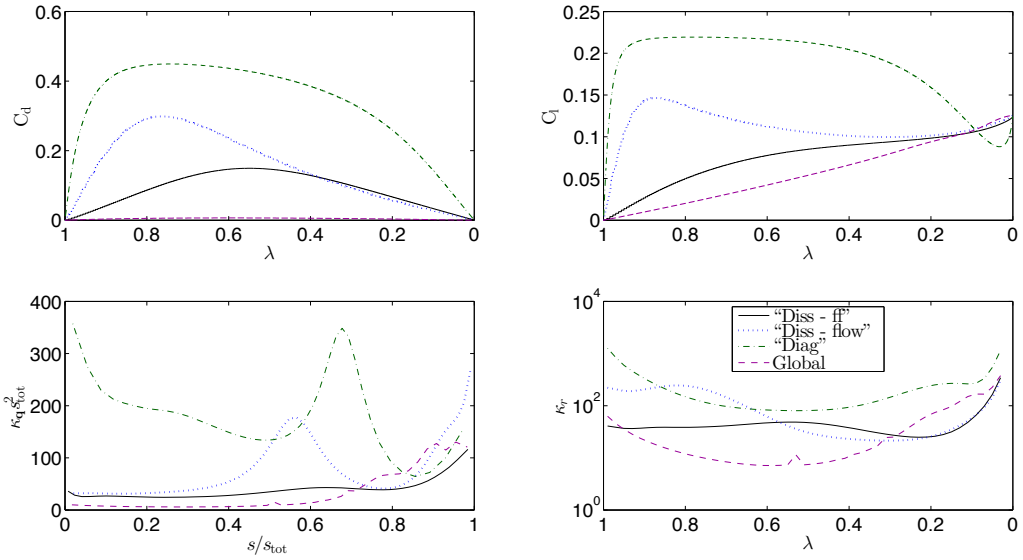


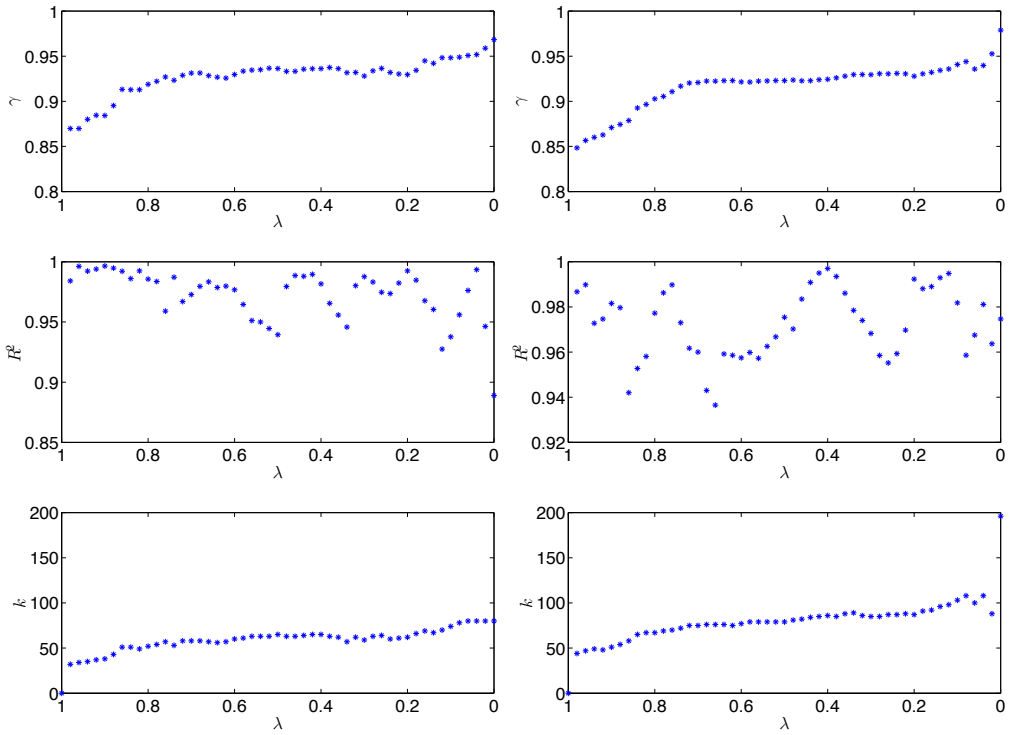
Figure 6: Surrogate curves and curvature profiles for several homotopies for inviscid flow over the NACA 0012 airfoil at Mach 0.3 and angle of attack of 1°

convex homotopy with the dissipation operator using both far-field and flow-imitative boundary conditions, the convex homotopy with the diagonal operator, and also for global homotopy. The surrogate curves are generated by accurately solving for points along the curve using a step size of $|\Delta\lambda| = 0.001$, whereas the curvature estimates use $|\Delta\lambda| = 0.02$ to avoid propagating too much error from the tangent calculation. The validation of the curvature calculation is performed by visual inspection of Figure A.19, from which we assess that the calculation is converged. The slight oscillations apparent for the global homotopy on the finest grid are most likely due to error propagation from the tangent calculation.

The surrogate curves and curvature profiles for this test case are shown in Figure 6. Clearly the lift and drag profiles do not accurately portray the curve traceability. The traceability of the different homotopy curves can be compared by considering the curvature profiles with respect to different parametrizations. Under both arclength and λ parametrizations, the curve generated by the convex homotopy with the diagonal operator appears to exhibit the lowest traceability, and the curve generated by the global homotopy appears to exhibit the highest traceability. Neither boundary condition type for the dissipation operator definitively stands out as giving a curve with better traceability over the other. A notable feature of the curvature profiles from the homotopies using the dissipation operators is the significant curvature increase near $s/s_{tot} = 1$, which indicates that these curves become increasingly difficult to trace near $\lambda = 0$.

Figure 7a shows the values of γ obtained by applying linear regression to the FGMRES convergence data using $\tau_l = 0.01$ according to equation (33). The profile for γ as a function of λ shows a similar trend to the number of iterations k as a function of λ, lending credibility to the metric. To show that γ is relatively independent of τ_l, the study is performed a second time using $\tau_l = 0.001$, as shown in Figure 7b.

To investigate the accuracy of the convergence model, the coefficients of determination (R^2



(a) $\tau_l = 0.01$ (b) $\tau_l = 0.001$

Figure 7: FGMRES convergence profile investigation for the convex homotopy with dissipation operator with far-field boundary conditions for inviscid flow over the NACA 0012 airfoil at Mach 0.3 and angle of attack of 1°

values) are plotted as a function of λ alongside the γ profiles in Figure 7. In addition, the convergence models are compared to the convergence data in Figure 8 at three values of λ. Comparing the plots in Figure 8 with the corresponding R^2 values in Figure 7a gives an idea of the relationship between the R^2 values and the quality of the fit. Clearly the model is not perfect and the metric should not be treated as more than a rough measure.

Values of γ vary between 0 and 1, with 0 being optimal and 1 indicating that the linear solver has stalled. However, it is not very representative of the cost of FGMRES and it does not provide an intuitive cost comparison. As apparent from equation (34), $1/\ln\gamma$ is proportional to k for a given linear solver residual target, making it a more representative metric of the actual cost of FGMRES than γ. When considering $1/\ln\gamma$, values near 0 indicate a high convergence rate and smaller (large magnitude negative) values indicate slower convergence. The FGMRES $1/\ln\gamma$ profiles are shown for the subsonic NACA 0012 test case for the various homotopies in Figure 9. It is apparent that the diagonal homotopy yields by far the best FGMRES convergence and that the global homotopy yields no significant improvement in convergence relative to the unmodified system.

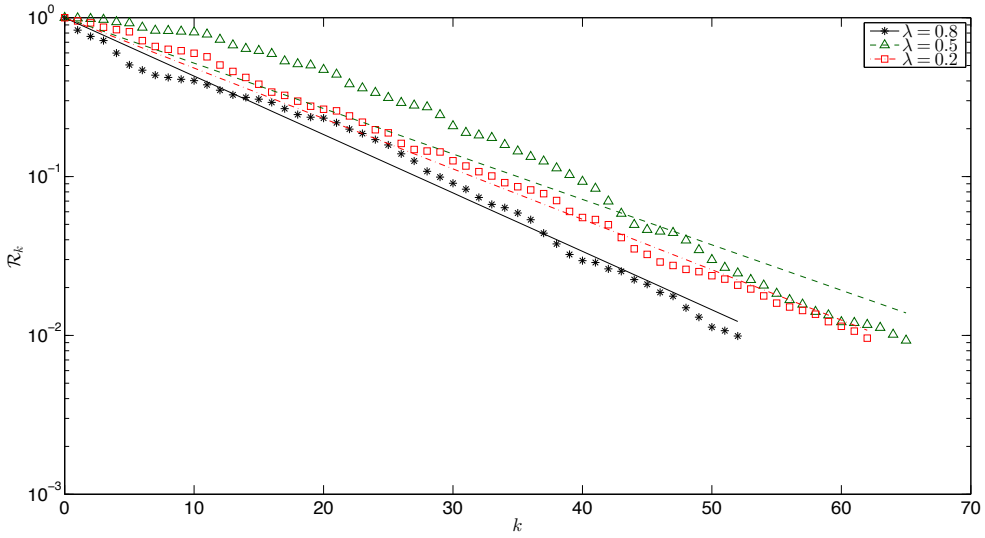


Figure 8: Comparison of FGMRES convergence profile model (solid lines) to convergence data ($*$) at $\tau_1 = 0.01$ for the convex homotopy with dissipation operator with far-field boundary conditions for inviscid flow over the NACA 0012 airfoil at Mach 0.3 and angle of attack of 1°

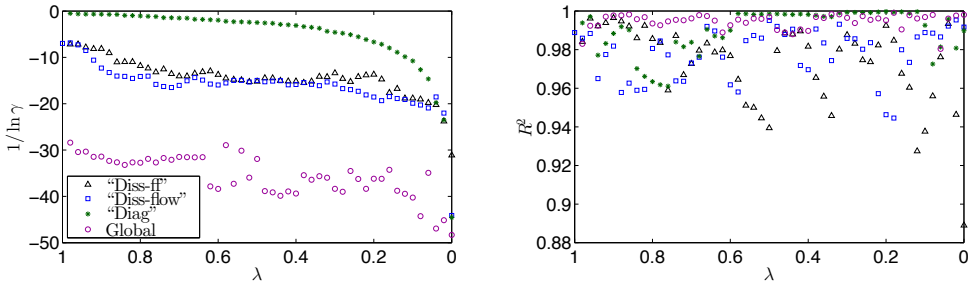


Figure 9: FGMRES convergence profiles for inviscid flow over the NACA 0012 airfoil at Mach 0.3 and angle of attack of 1°

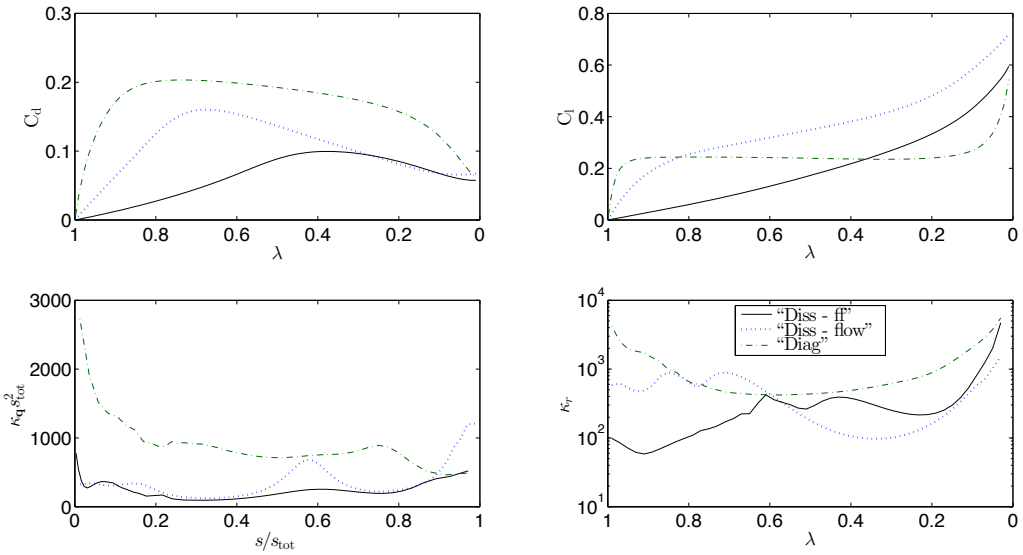


Figure 10: Surrogate curves and curvature profiles for several homotopies for inviscid flow over the NACA 0012 airfoil at Mach 0.8 and angle of attack of 3°

7.3. Curvature Profiles for a Two-Dimensional Transonic Inviscid Flow

This study is analogous to the previous one except that it is performed at the transonic conditions of Mach 0.8 and angle of attack 3°. Validation of the curvature calculations is shown in Figure A.20. The surrogate curves and curvature profiles are shown in Figure 10.

The global homotopy case could not be converged, hence data is not shown for the global homotopy case. For the other three homotopies, the C_l and C_d surrogate curves profiles show surprising similarity to the surrogate curves from the subsonic example. As with the subsonic example, the traceability of the convex homotopy with the diagonal operator is lower than the traceability using the dissipation operator, with the far-field boundary conditions showing small improvement over flow-imitative boundary conditions.

7.4. A Demonstration of the Effects of μ-Scaling

As an example illustrating the effect of μ-scaling, the C_l and C_d surrogate curves, as well as the curvature profiles, are shown in Figure 11 for the inviscid NACA 0012 test case at Mach number 0.3 and angle of attack 1°. The convex homotopy with the dissipation operator and far-field boundary conditions is used. It can be seen from the C_l and C_d surrogate curves that setting $\mu_u > 1$ has the effect of stretching the curve near $\lambda = 1$ and compressing it near $\lambda = 0$, while choosing $\mu_u < 1$ has the reverse effect. This is also reflected by the s vs. λ plot, where it can be seen that more of the arclength s is traversed early on when using smaller μ_u, or by the $\kappa_{\mathbf{q}}$ vs. λ plot, which shows how the partial curvature $\kappa_{\mathbf{q}}$ has been redistributed.

The following plot, which shows $\kappa_{\mathbf{q}} s_{tot}^2$ vs. s/s_{tot}, shows much less dramatic redistribution of the curvature, indicating that the value of μ has much less impact on the performance of a continuation method which takes constant Δs than a continuation method which takes constant $\Delta\lambda$. The plot of κ_r vs. λ provides a metric for how difficult it would be to trace each of the three

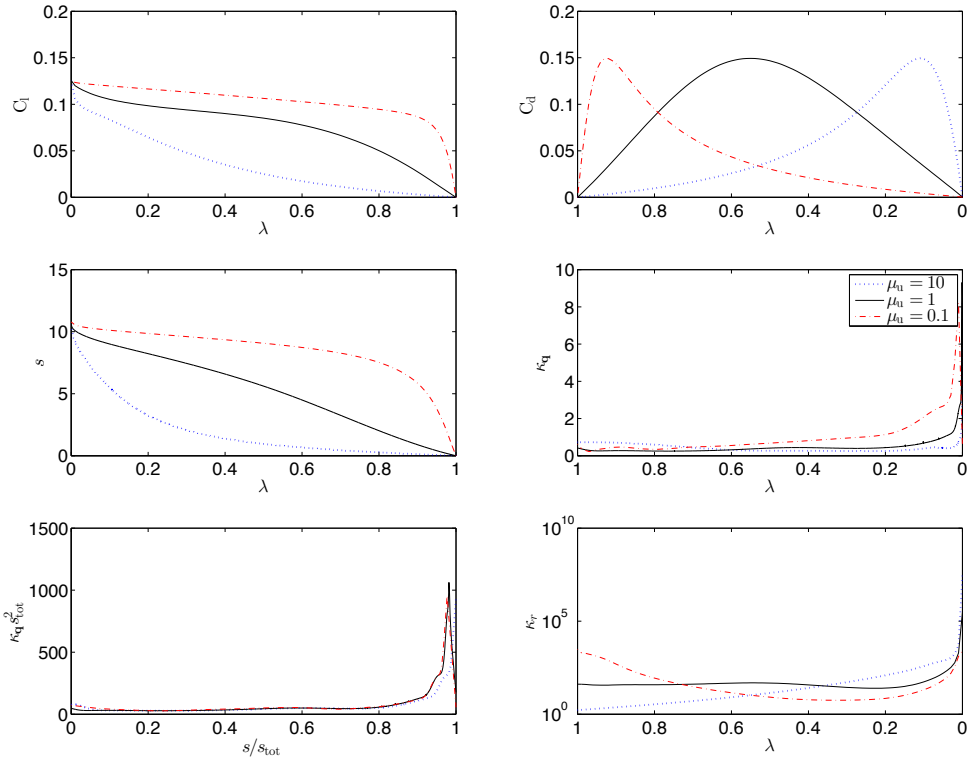


Figure 11: Effect of μ_u on the homotopy using the dissipation operator with far-field boundary conditions for the inviscid NACA 0012 test case at Mach 0.3 and angle of attack 1°

curves if constant $\Delta\lambda$ is used. Ideally, μ is chosen to make the κ_r profile is as flat as possible, indicating even curvature distribution. It is apparent from the bottom right subplot that $\mu_u = 1$ provides the most consistent κ_r profile throughout traversing. This indicates that the value of $\mu_a = 0.7$, which, as discussed in Section 2.8, was obtained previously [4] from a numerical algorithm applied to a different case, is effective for this case as well.

7.5. Curvature Profiles for a Three-Dimensional Inviscid Flow and Effect of Mesh Refinement

The purpose of this study is two-fold: to profile the homotopies for three-dimensional inviscid flows and to investigate the effects of grid refinement on the homotopy. For global homotopy, the homotopy residual has a continuous counterpart which is obtained by taking the limit of zero mesh spacing. As such, the homotopy itself is expected to converge to a grid-converged value in the limit of zero mesh spacing in much the same way as the flow solution does. Furthermore, since the flow residual is second-order accurate, the homotopy should also be second-order accurate in the same sense as the flow residual; that is, the error in the homotopy on a current mesh calculated relative to its continuous counterpart is expected to be proportional to the grid spacing squared.

For convex homotopy, the homotopy will only have a continuous counterpart if the homotopy system is mesh-independent. Since both the dissipation operator and diagonal operator presented

in this paper do not have mesh-independent continuous counterparts, the deformation can change in a more fundamental way as the mesh is refined, even if the mesh is already fine enough to give a very accurate flow solution.

The surrogate curves and curvature profiles are generated for inviscid flow over the ONERA M6 wing at Mach 0.4 and angle of attack 3°. The grid consists of 1.9208×10^6 nodes divided evenly into 32 blocks. The fine version of the mesh is generated by doubling the number of nodes in each direction, the location of the new nodes determined by interpolation from a B-spline parametrization of the grid [17, 47], which is important in order to fit the ONERA M6 wing smoothly on the fine mesh. Each block is then split evenly into 8 blocks, resulting in 256 blocks in total. Validation of the curvature calculation is shown in Figures A.21 and A.22 for the two meshes. The λ spacing appears to be sufficiently small for all cases, though the error is somewhat significant for the global homotopy on the fine mesh.

Since there are 8 times more points on the fine mesh than the original, the $\kappa_{\mathbf{q}} s_{\text{tot}}^2$ and κ_r values will naturally increase by a factor of $\sqrt{8}$. The reason for this can be seen from the curvature equations. For example, if instead the nodes had increased by a factor of two (consider, for example, the 1D case), then the additional elements appended to the $\mathbf{q}$ vector are very close in size to the original values and so $\kappa_{r,\text{fine}} \sim \sqrt{\dot{\mathbf{q}}(r) \cdot \dot{\mathbf{q}}(r) + \dot{\mathbf{q}}(r) \cdot \dot{\mathbf{q}}(r)} = \sqrt{2}\kappa_r$. This factor is simply due to having more state elements in the approximation to the continuous deformation and does not indicate decreased traceability on the finer mesh. To account for this effect, $\kappa_{\mathbf{q}} s_{\text{tot}}^2 / \sqrt{N}$ and $\kappa_r / \sqrt{N}$ are used as traceability metrics instead of the usual $\kappa_{\mathbf{q}} s_{\text{tot}}^2$ and κ_r. Note that this modification to the traceability metrics is only applicable to uniform mesh refinement and is not suitable for comparing traceability on different meshes in general.

Figure 12 shows the comparison of the surrogate curves and curvature profiles on the original mesh and finer mesh. The C_L and C_D profiles appear similar on both meshes for all convex homotopies and unchanged for the global homotopy. Traceability, when including $1/\sqrt{N}$ in the metric, has been affected noticeably but not dramatically by the mesh refinement for all homotopies.

The FGMRES convergence profiles are also shown for both meshes in Figure 13. On the coarse mesh, the linear solver converges fastest under the convex homotopy with the diagonal operator, followed by the dissipation operators, followed by global homotopy, which is consistent with performance for the inviscid subsonic NACA 0012 case. When the mesh is refined, the convergence rate of the linear solver becomes universally slower for all homotopies and for all $\lambda \in \Lambda$. Oddly, on the fine mesh the dissipation operators seem to actually worsen the linear solver performance relative to the linear systems associated with $\mathcal{R}(\mathbf{q})$.

7.6. Curvature Profiles for a Three-Dimensional Laminar Flow and Effect of Grid Topology

This test case is laminar flow over the ONERA M6 wing at Mach 0.3, angle of attack 1°, and Reynolds number 1×10^3. Two grids are used: the first grid consists of 2.11×10^6 nodes divided evenly into 48 blocks and the second grid consists of 1.88×10^6 nodes divided evenly into 16 blocks. Though the grids are of similar size and refinement, the first one has an H-H topology whereas the second one has an H-C topology. Validation of the curvature estimates is shown in Figures A.23 and A.24 for the two grids.

The homotopies on the two grids are compared in Figure 14. The surrogate curves and curvature profiles appear to be quite similar on the two grids for the three convex homotopies investigated. In addition, FGMRES convergence profiles are shown in Figure 15. The R^2 values

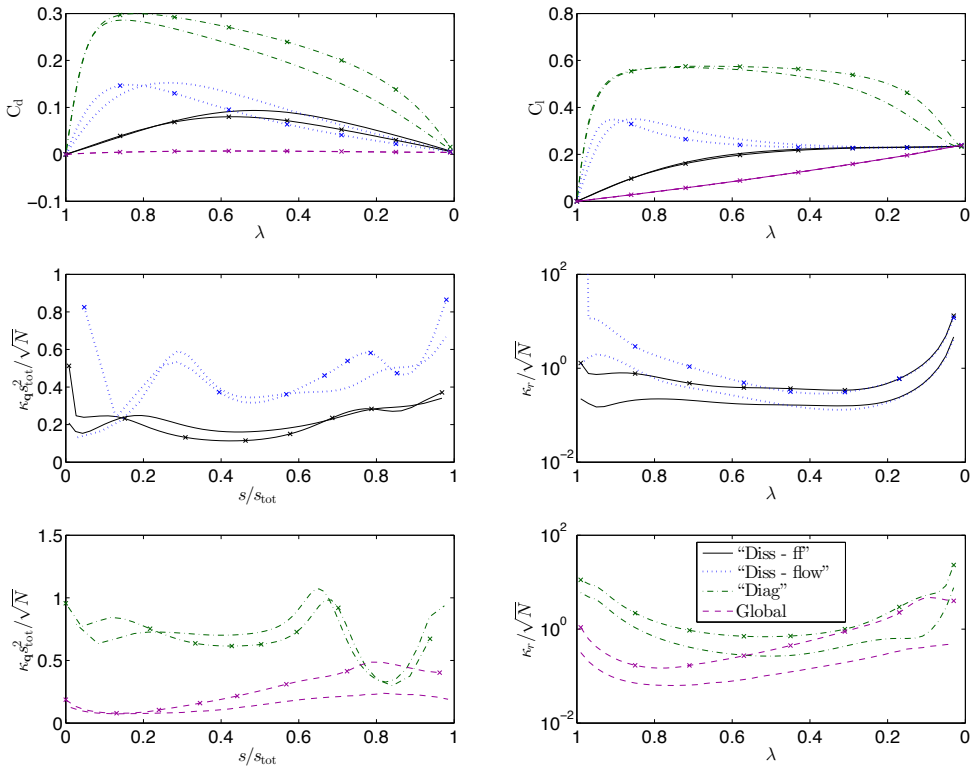
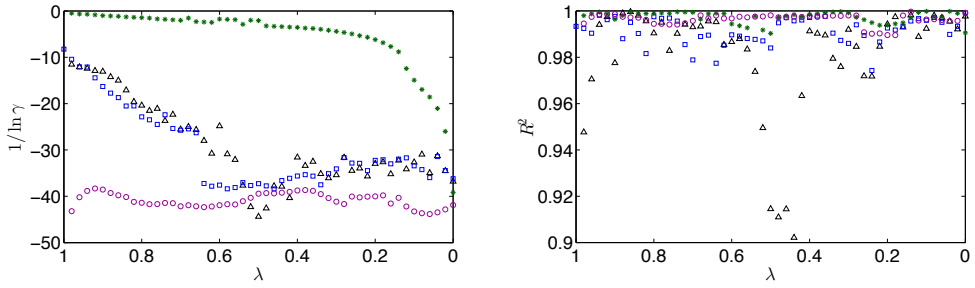
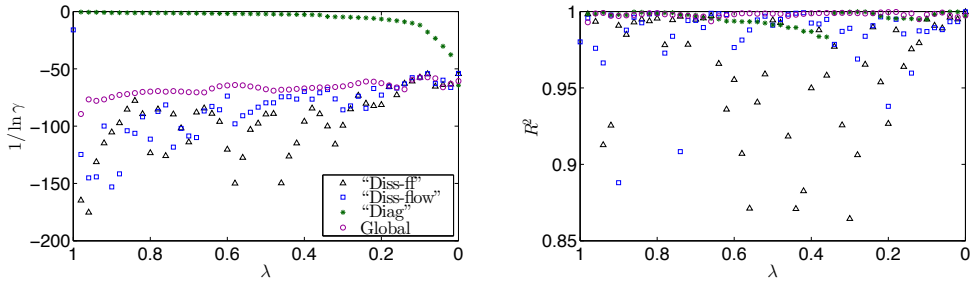


Figure 12: Curvature profiles and the effects of mesh refinement on several homotopies for inviscid flow over the ONERA M6 geometry at Mach 0.4 and angle of attack of 3°; the lines marked with x's represent the data for the finer grid



(a) Coarse mesh



(b) Fine mesh

Figure 13: FGMRES convergence profiles for inviscid flow over the ONERA M6 geometry at Mach 0.4 and angle of attack of 3°

for this case indicate less agreement of the linear residual data with the convergence model assumed. This is especially true for the dissipation operator with far-field boundary conditions on the H-C topology mesh, which may be disregarded entirely. Data for the global homotopy is not shown because we were unable to converge this case using the global homotopy.

7.7. Curvature Profiles for a Two-Dimensional Turbulent Flow

This study is a curvature profiling for turbulent flow over the NACA 0012 airfoil using the RANS-SA equations. The grid consists of 19200 nodes divided evenly into 8 blocks. The Reynolds number based on the chord length for the test case is 4×10^6, the Mach number is 0.4, and the angle of attack is 1°. Validation of the curvature estimates is shown in Figure A.25. The diagonal operator, in this case, appears to be somewhat under-converged, while at the same time exhibiting oscillations indicating that $|\Delta\lambda|$ is sufficiently small such that error propagated from the tangent estimation is becoming significant. It appears also that some features may be missing from the κ_r profiles near $\lambda = 1$ for all homotopies due to insufficient resolution in $\Delta\lambda$.

The curvature profiles are shown in Figure 16 and are more complicated in this case. The s/s_{tot} vs. λ plot is helpful to put the curvature profiles into the proper context. Most of the work is done once λ becomes small. For the diagonal case, at $\lambda = 0.05$, s/s_{tot} is still only 0.04, indicating that only 4% of the curve has been traversed, measured in distance along the curve. For small λ, the curvature would thus appear much lower moving along the curve than to an outside observer using the λ coordinate system, as the curvature has been stretched out along the length of the curve.

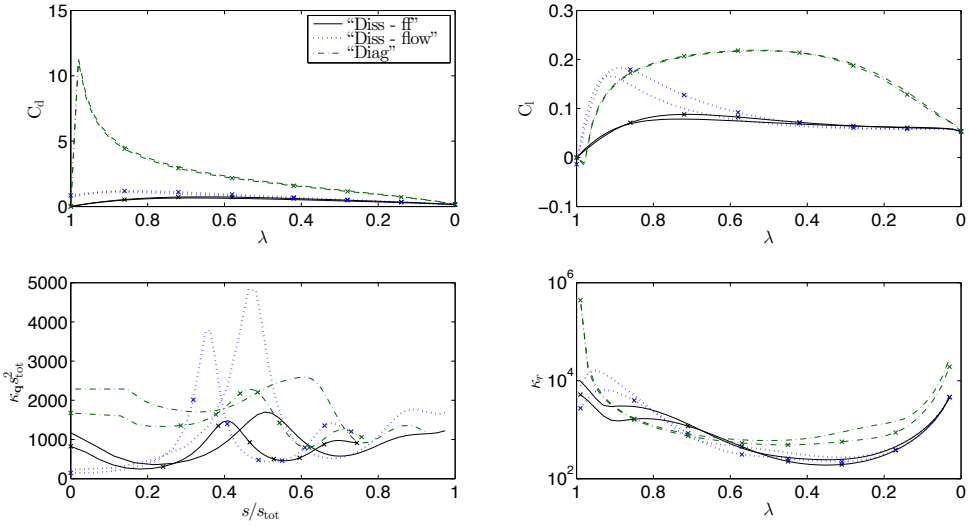
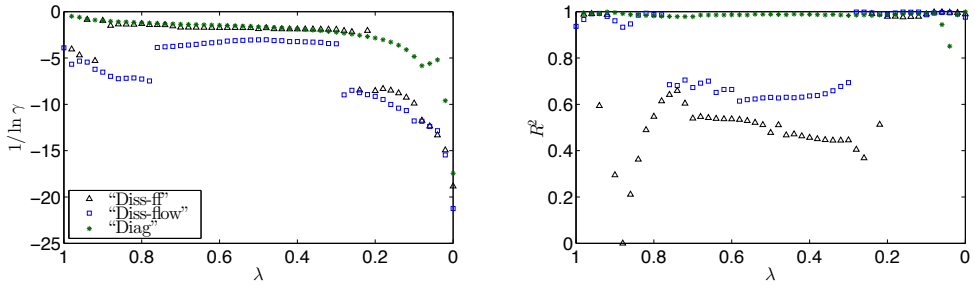
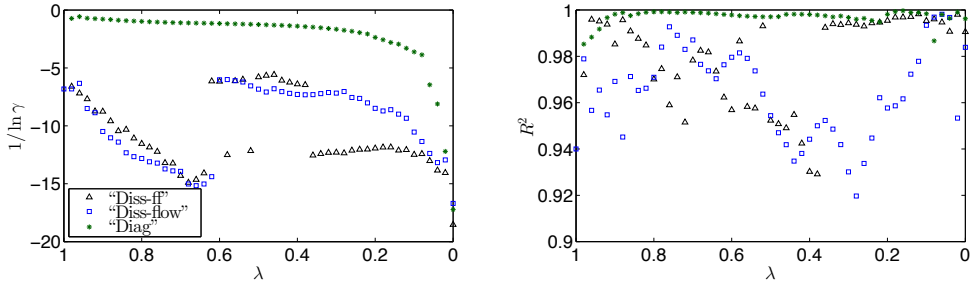


Figure 14: Curvature profiles and the effects of grid topology on several homotopies for laminar flow over the ONERA M6 geometry at Mach 0.4, angle of attack 3°, and Reynolds number 1×10^3; the lines marked with x's represent the data for the H-H grid



(a) H-C topology mesh



(b) H-H topology mesh

Figure 15: FGMRES convergence profiles for laminar flow over the ONERA M6 geometry at Mach 0.4, angle of attack 3°, and Reynolds number 1×10^3

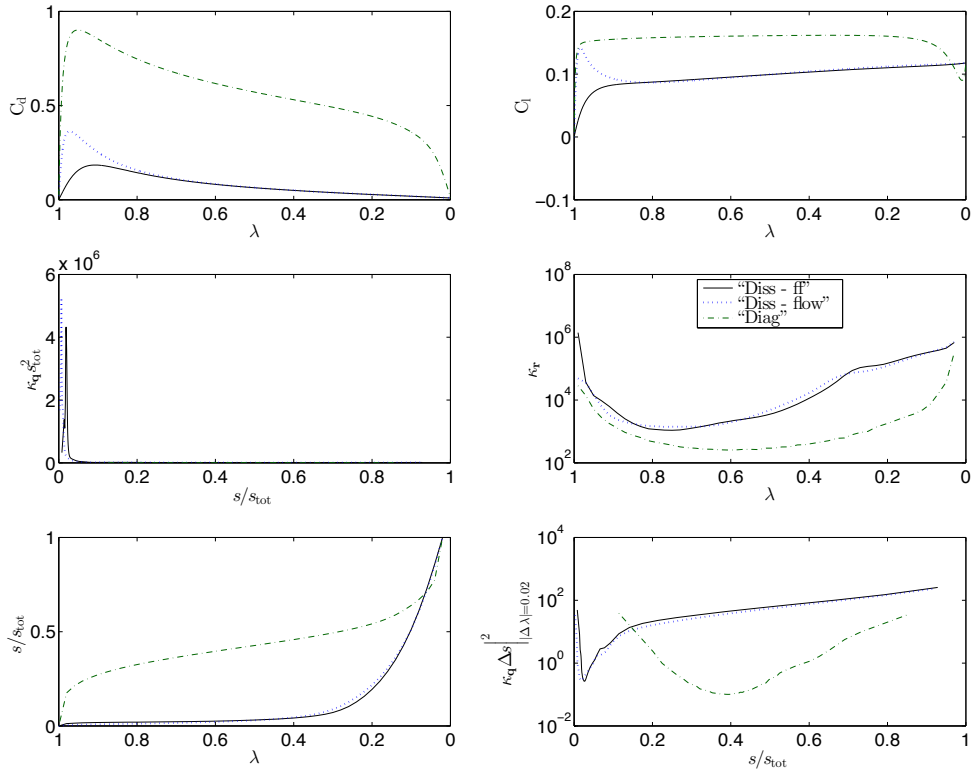


Figure 16: Curvature profiles of several homotopies for turbulent flow over the NACA 0012 airfoil at Mach 0.4, Reynolds number 4×10^6, and angle of attack of 1°

To give a better indication of the local curve traceability, the analysis is also performed by plotting $\kappa_{\mathbf{q}} \Delta s|^2_{|\Delta\lambda|=0.02}$ against s/s_{tot}, from which a very different profile is apparent. This metric is different from κ_r; it still represents the traceability assuming constant Δs except that the actual value of Δs assumed is that locally associated with $|\Delta\lambda| = 0.02$. The traceability using this metric may be a more representative metric of the performance of a continuation algorithm and appears quite different from the usual $\kappa_{\mathbf{q}} s_{\text{tot}}$ metric. It is clear from the plots that the curvature distribution is very imbalanced for this homotopy. This issue cannot be addressed by a simple change of coordinates.

7.8. Curvature Profile for a Three-Dimensional Turbulent Flow

This case is a three-dimensional turbulent flow over the ONERA M6 wing. The study is performed using an H-H topology mesh consisting of 2.33×10^6 nodes divided evenly into 192 blocks. The flow conditions are Mach 0.4, angle of attack 3°, and Reynolds number 1×10^6. The validation attempt is shown in Figure A.26. Due to significant error apparent in the validation attempt, the curvature profiles are not shown for this case. The C_L and C_D profiles are shown in Figure 18. What is remarkable about the profiles is the clear similarity to the C_l and C_d profiles obtained for the two-dimensional NACA 0012 case shown in Figure 16.

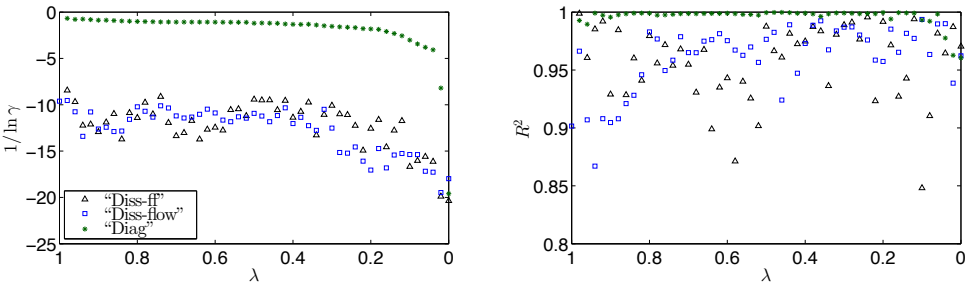


Figure 17: FGMRES convergence profiles for turbulent flow over the NACA 0012 geometry at Mach 0.4, Reynolds number 4×10^6, and angle of attack of 1°

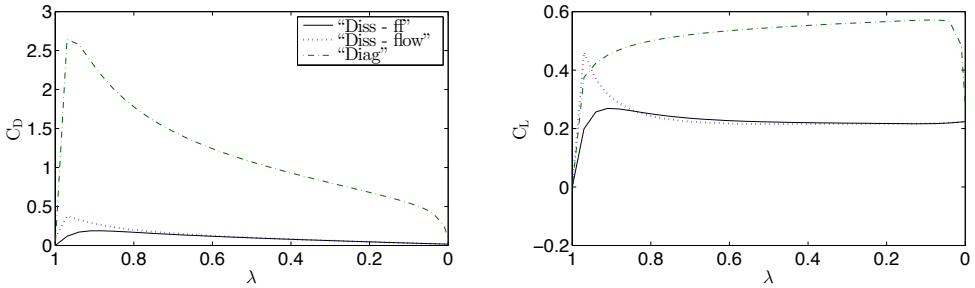


Figure 18: C_L and C_D profiles of several homotopies for turbulent flow over the ONERA M6 geometry at Mach 0.4, Reynolds number 10^6, and angle of attack of 3°

8. Summary of Observations Based on the Results

The cases studied were one-dimensional inviscid subsonic and transonic flow through a converging-diverging nozzle, two-dimensional inviscid subsonic and transonic flows over the NACA 0012 airfoil, three-dimensional subsonic laminar flow over the ONERA M6 wing, two-dimensional turbulent subsonic flow over the NACA 0012 airfoil, and three-dimensional turbulent subsonic flow over the ONERA M6 wing. A summary is provided based on the studies presented on these cases.

8.1. Global Homotopy

For subsonic inviscid cases, this homotopy exhibits the highest traceability but the augmented operator under this homotopy offers no benefit to linear solver convergence compared to the unmodified residual. We were unable to converge any inviscid transonic cases with the global homotopy except for a one-dimensional case, which required extremely conservative parameter settings. The surrogate curves in this case show a sharp and dramatic change in curve values with respect to λ at the point where the shock appeared in the solution. This may indicate why the two- and three-dimensional cases failed. We were also unable to converge any of the viscous cases using this homotopy.

8.2. Convex Homotopy – Diagonal Operator

The diagonal homotopy was found to give the greatest benefit to linear system performance but generally exhibited the highest curvature. The expected effect, from a continuation perspective, is a greater number of nonlinear iterations but each coming at less cost than a continuation algorithm based on any of the other homotopies investigated. However, it tends to be less robust, especially for RANS cases, than the dissipation operators, due to the reduced traceability, and we were unable to converge some of the cases under consideration.

8.3. Convex Homotopy – Dissipation Operators

While not offering much improvement to the linear solver performance, and actually worsening performance on one of the fine mesh cases investigated, the homotopies based on dissipation operators exhibit modest curvature compared to the diagonal operator and are more versatile than the global homotopy in that we were able to converge all cases under consideration with these operators. Of the two choices of boundary conditions under consideration, "far-field" and "flow-imitative", traceability and linear solver performance were both quite similar, but traceability using "far-field" boundary conditions was generally marginally better.

9. Conclusions

A framework for assessing the suitability of homotopies for continuation has been presented and demonstrated through application to several specific homotopies for application to computational aerodynamics problems. Identifying and understanding the underlying features which affect performance of homotopy continuation algorithms is important in designing efficient and robust homotopy continuation algorithms. The knowledge gained of the specific homotopies presented in this paper has been summarized in Section 8.

The effects of mesh refinement and changes in mesh topology were quantified using some examples, demonstrating that traceability is, as predicted, relatively insensitive to these effects.

This is because the homotopy curve is not fundamentally altered by refining the mesh or changing the mesh topology. However, the cost of the linear solver is generally expected to grow super-linearly with mesh size, so that algorithm cost is still a super-linear function of mesh size. This knowledge is useful in selecting continuation parameters such as step size, indicating that these parameters should not change with mesh refinement. This contrasts with pseudo-transient continuation where it is well-known that more conservative parameters are generally needed on finer meshes. If the homotopy algorithm exhibits reduced performance when the mesh is refined, it is most likely the linear solver parameters which should be adjusted.

The practical effect of re-parametrizing the curve with respect to its continuation parameter λ was investigated numerically. While this re-parametrization does not affect the implicitly-defined curve, it does in practice affect the performance of a practical continuation algorithm which likely includes restrictions on the minimum and maximum allowable $|\Delta\lambda|$ and perhaps the ratio of consecutive $|\Delta\lambda|$. While re-parametrizing the curve can make it more amenable to continuation, identifying the optimal parametrization requires numerical experimentation.

Acknowledgments

The authors gratefully acknowledge financial assistance from the Natural Science and Engineering Research Council (NSERC), the Canada Research Chairs program, and the University of Toronto. Computations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto.

References

[1] Allgower, E. L., Georg, K., 1990. Introduction to Numerical Continuation Methods. Society for Industrial and Applied Mathematics.

[2] Bates, D. J., Hauenstein, J. D., Sommese, A. J., 2011. Efficient path tracking methods. Numer. Algorithms 58 (4), 451–459.

[3] Brown, D. A., 2015. Efficient homotopy continuation algorithms with application to computational fluid dynamics. Ph.D. thesis, University of Toronto, Toronto, Ontario, Canada.

[4] Brown, D. A., Zingg, D. W., June 2013. Advances in homotopy-based globalization strategies in computational fluid dynamics. AIAA-2013-2944.

[5] Brown, D. A., Zingg, D. W., 2016. Efficient numerical differentiation of implicitly-defined curves for sparse systems. J. Comput. Appl. Math. 304, 138–159.

[6] Brown, D. A., Zingg, D. W., 2016. A monolithic homotopy continuation algorithm with application to computational fluid dynamics. J. Comp. Phys. 321, 55–75.

[7] Broyden, C. G., 1965. A class of methods for solving nonlinear simultaneous equations. Math. Comp. 19, 577–593.

[8] Carpenter, M. H., Gottlieb, D., Abarbanel, S., 1994. Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes. J. Comput. Phys. 111 (2), 220–236.

[9] Del Rey Fernández, D. C., Hicken, J. E., Zingg, D. W., 2014. Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations. Comput. Fluids 95, 171–196.

[10] Embree, M., 1999. How descriptive are GMRES convergence bounds? Tech. rep., Oxford University Computing Laboratory.

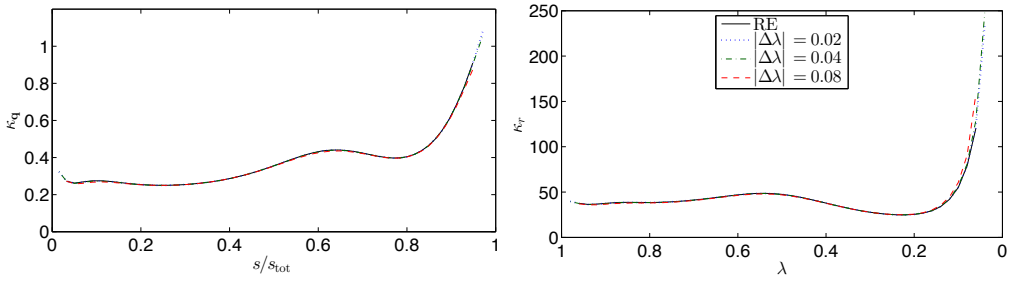
[11] Funaro, D., Gottlieb, D., 1988. A new method of imposing boundary conditions in pseudospectral approximations of hyperbolic equations. Math. Comput. 51, 599–613.

[12] Greenbaum, A., Pták, V., Strakoš, A., 1996. Any nonincreasing convergence curve is possible for GMRES. SIAM J. Matrix Anal. Appl. 17 (3), 465–469.

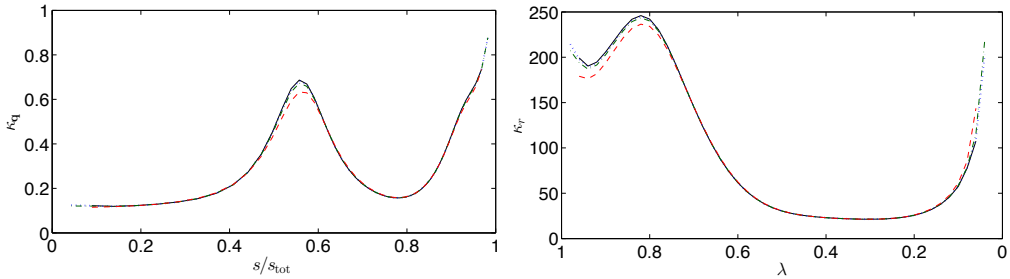
[13] Hao, W., Hauenstein, J. D., Shu, C., Sommese, A. J., Xu, Z., Zhang, Y., 2013. A homotopy method based on WENO schemes for solving steady state problems of hyperbolic conservation laws. J. Comput. Phys. 250, 332–346.
[14] Hicken, J. E., Buckley, H., Osusky, M., Zingg, D. W., June 2011. Dissipation-based continuation: a globalization for inexact-Newton solvers. AIAA 2011-3237.
[15] Hicken, J. E., Zingg, D. W., 2008. A parallel Newton-Krylov solver for the Euler equations discretized using simultaneous approximation terms. AIAA J. 46 (11), 2773–2786.
[16] Hicken, J. E., Zingg, D. W., June 2009. Globalization strategies for inexact-Newton solvers. AIAA-2009-4139.
[17] Hicken, J. E., Zingg, D. W., 2010. Aerodynamic optimization algorithm with integrated geometry parameterization and mesh movement. AIAA J. 48 (2), 401–413.
[18] Hirsch, C., 1988. Numerical Computation of Internal and External Flows. John Wiley & Sons.
[19] Jameson, A., Schmidt, W., Turkel, E., June 1981. Numerical solution of the Euler equations by finite-volume methods using Runge-Kutta time-stepping schemes. AIAA-1981-1259.
[20] Knoll, D. A., Keyes, D. E., 2004. Jacobian-free Newton-Krylov methods: a survey of approaches and applications. J. Comput. Phys. 193, 357–397.
[21] Kreiss, H., Scherer, G., 1974. Finite element and finite difference methods for hyperbolic partial differential equations. In: de Boor, C. (Ed.), Mathematical Aspects of Finite Elements in Partial Differential Equations: proceedings of a symposium conducted by the Mathematics Research Center, the University of Wisconsin. Mathematics Research Center, the University of Wisconsin, Academic Press, pp. 195–212.
[22] Kreyszig, E., 1959. Differential Geometry. University of Toronto Press, Toronto, Ontario, Canada.
[23] Lomax, H., Pulliam, T. H., Zingg, D. W., 2001. Fundamentals of Computational Fluid Dynamics. Springer-Verlag.
[24] Lundberg, B. N., Poore, A. B., 1991. Variable order Adams-Bashforth predictors with an error stepsize control for continuation methods. SIAM J. Sci. Stat. Comp. 12 (3), 695–723.
[25] Mackens, W., 1989. Numerical differentiation of implicitly defined space curves. Computing 41, 237–260.
[26] Meyer, G. H., 1968. On solving nonlinear equations with a one-parameter operator embedding. SIAM J. Numer. Anal. 5 (4), 739–752.
[27] Morgan, A., Sommese, A., 1987. A homotopy for solving general polynomial systems that respect m-homogeneous structures. Appl. Math. Comput. 24 (2), 101–113.
[28] Nielsen, E. J., Anderson, W. K., Walters, R. W., Keyes, D. E., June 1995. Application of Newton-Krylov methodology to a three-dimensional unstructured Euler code. AIAA-95-1733.
[29] Osusky, M., Zingg, D. W., 2013. A parallel Newton-Krylov-Schur flow solver for the Navier-Stokes equations discretized using summation-by-parts operators. AIAA J. 51 (12), 2833–2851.
[30] Pönisch, G., Schwetlick, H., 1981. Computing turning points of curves implicitly defined by nonlinear equations depending on a parameter. Computing 26, 107–121.
[31] Pulliam, T. H., 1986. Artificial dissipation models for the Euler equations. AIAA J. 24 (12), 1931–1940.
[32] Pulliam, T. H., Zingg, D. W., 2014. Fundamental Algorithms in Computational Fluid Dynamics. Springer.
[33] Riley, D. S., Winters, K. H., 1990. A numerical bifurcation of natural convection in a tilted two-dimensional porous cavity. J. Fluid Mech. 215, 309–329.
[34] Saad, Y., 2003. Iterative Methods for Sparse Linear Systems, 2nd Edition. SIAM, Philadelphia, PA.
[35] Saad, Y., Schultz, M., 1986. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comp. 7, 856–869.
[36] Sanchez, J., Marques, F., Lopez, J. M., 2002. A continuation and bifurcation technique for Navier-Stokes flows. J. Comput. Phys. 180, 78–98.
[37] Schmitt, V., Charpin, F., May 1979. Pressure distributions on the ONERA-M6-Wing at transonic Mach numbers. Tech. rep., Experimental Data Base for Computer Program Assessment, report of the Fluid Dynamics Panel Working Group 04, AGARD AR 138.
[38] Schwetlick, H., Cleve, J., 1987. Higher order predictors and adaptive steplength control in path following algorithms. SIAM J. Numer. Anal. 24 (6), 1382–1393.
[39] Sommese, A. J., Wampler, C. W., 2005. The Numerical Solution of Systems of Polynomials Arising in Engineering and Science. World Publishing Company.
[40] Spalart, P. R., Allmaras, S. R., January 1992. A one-equation turbulence model for aerodynamic flows. AIAA 92-0439.
[41] Syam, M. I., Siyyam, H. I., 1999. Numerical differentiation of implicitly defined curves. J. Comput. Appl. Math. 108, 131–144.
[42] Titley-Peloquin, D., Pestana, J., Wathen, A. J., 2014. GMRES convergence bounds that depend on the right-hand-side vector. IMA J. Numer. Anal. 34 (2), 462–479.
[43] Wagner, C. F., 2015. Improving shock-capturing robustness for higher-order finite element solvers. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States.
[44] Wales, C., Gaitonde, A. L., Jones, D. P., Avitabile, D., Champneys, A. R., 2012. Numerical continuation of high Reynolds number external flows. Int. J. Numer. Meth. Fl. 68 (2), 135–159.

[45] Winters, K. H., Cliffe, K. A., 1985. The prediction of critical points for thermal explosions in a finite volume. Combust. Flame 62 (1), 13–20.
[46] Yu, M., Wang, Z. J., 2016. Homotopy continuation of the high-order flux reconstruction/correction procedure via reconstruction (FR/CPR) method for steady flow simulation. Comput. Fluids 131, 16–28.
[47] Yu, T.-Y., Soni, B. K., 1995. Application of NURBS in numerical grid generation. Comput. Aided Design 27, 147–157.

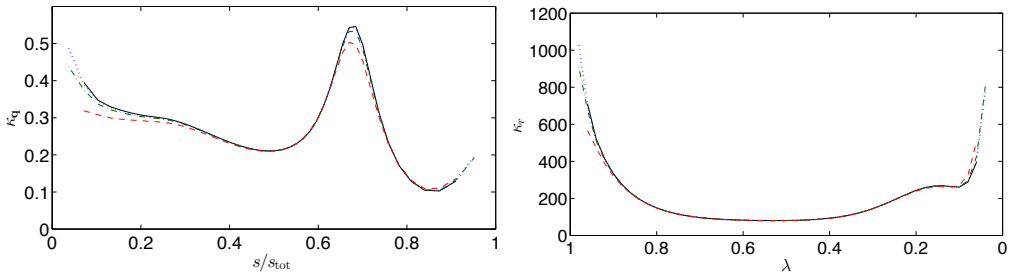
Appendix A. Additional Figures



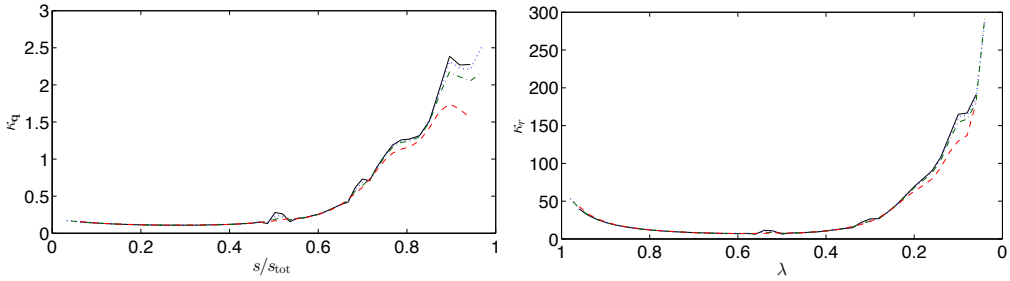
(a) Convex homotopy with dissipation operator with far-field boundary conditions



(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

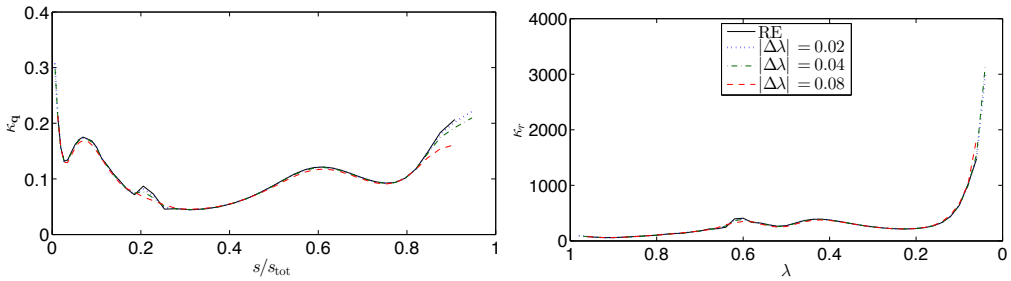


(c) Convex homotopy with diagonal operator

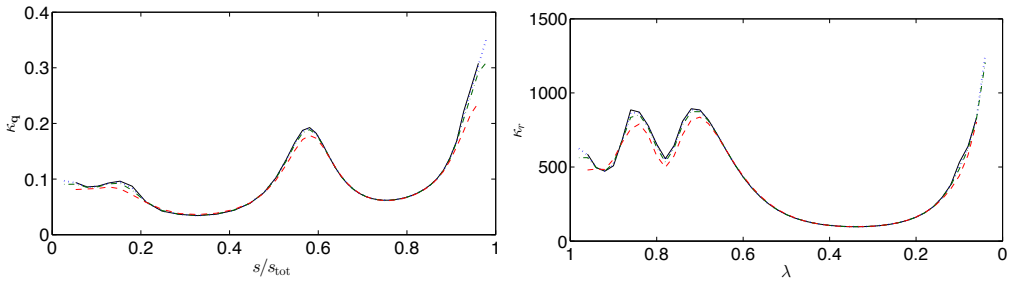


(d) Global homotopy

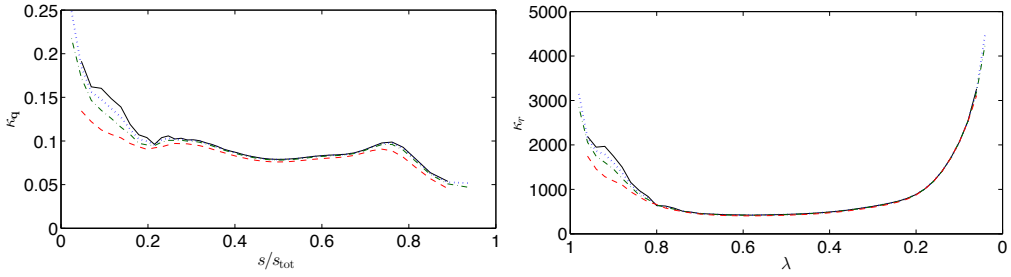
Figure A.19: Validation of the curvature estimation for inviscid flow over the NACA 0012 airfoil at Mach 0.3 and angle of attack of 1°



(a) Convex homotopy with dissipation operator with far-field boundary conditions

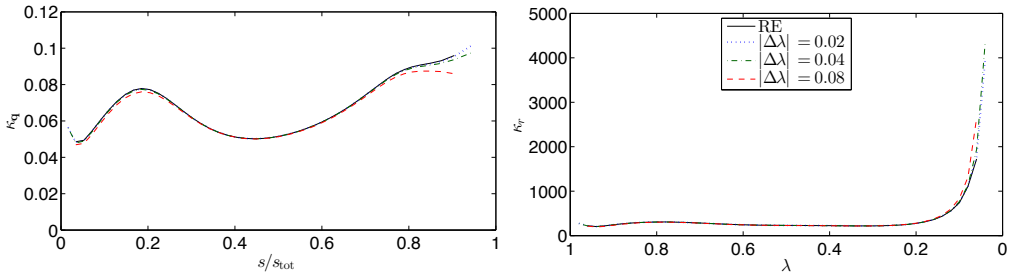


(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

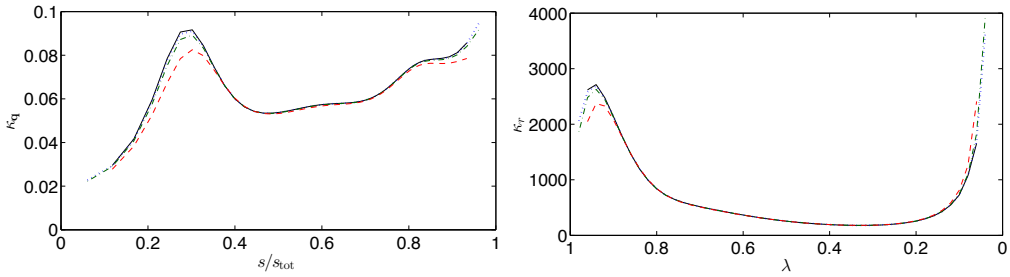


(c) Convex homotopy with diagonal operator

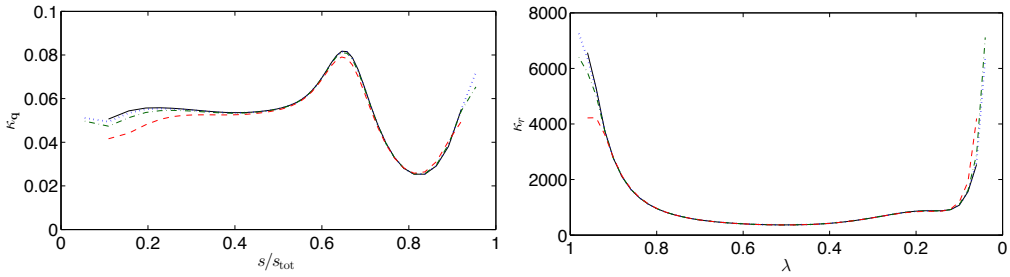
Figure A.20: Validation of the curvature estimation for inviscid flow over the NACA 0012 airfoil at Mach 0.8 and angle of attack of 3°



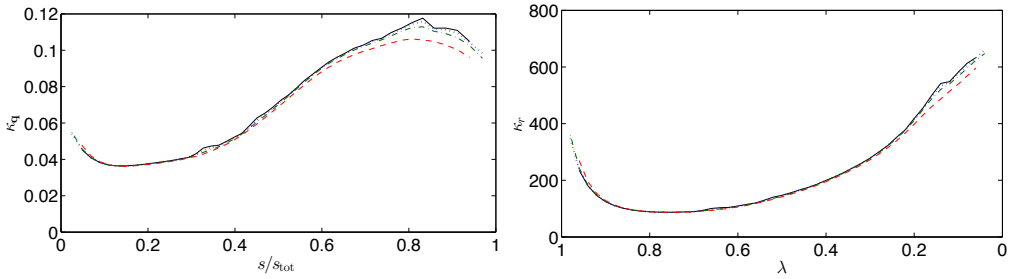
(a) Convex homotopy with dissipation operator with far-field boundary conditions



(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

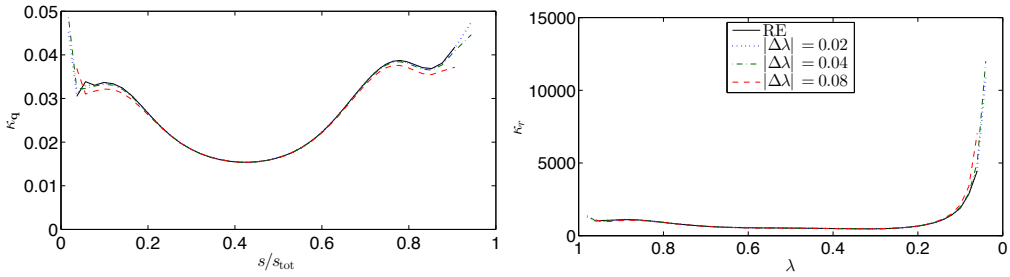


(c) Convex homotopy with diagonal operator

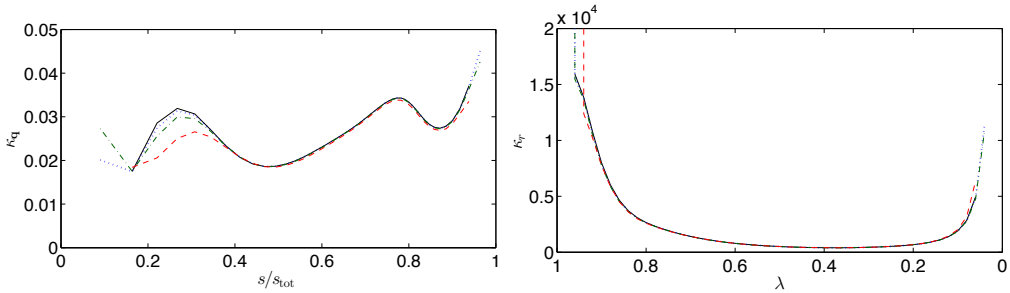


(d) Global homotopy

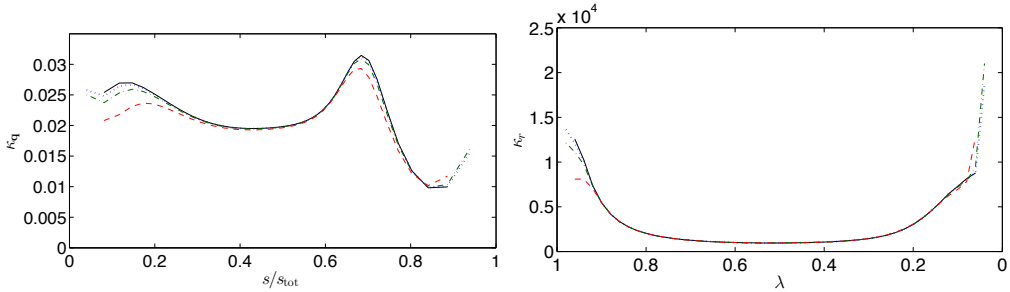
Figure A.21: Validation of the curvature estimation for inviscid flow over the ONERA M6 geometry on the coarse mesh at Mach 0.4 and angle of attack of 3°



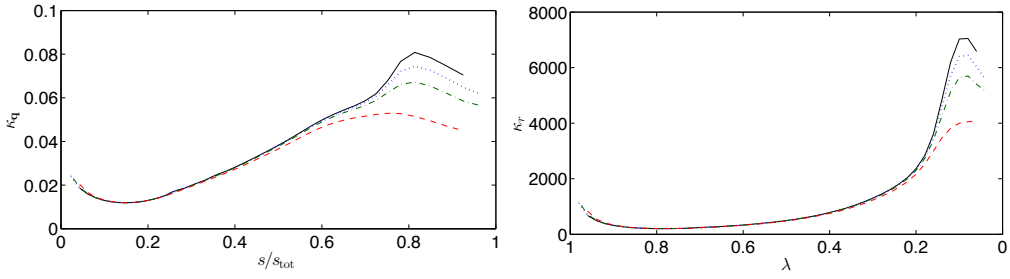
(a) Convex homotopy with dissipation operator with far-field boundary conditions



(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

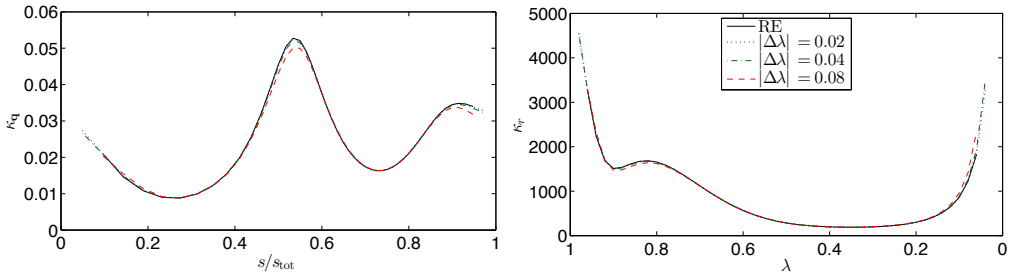


(c) Convex homotopy with diagonal operator

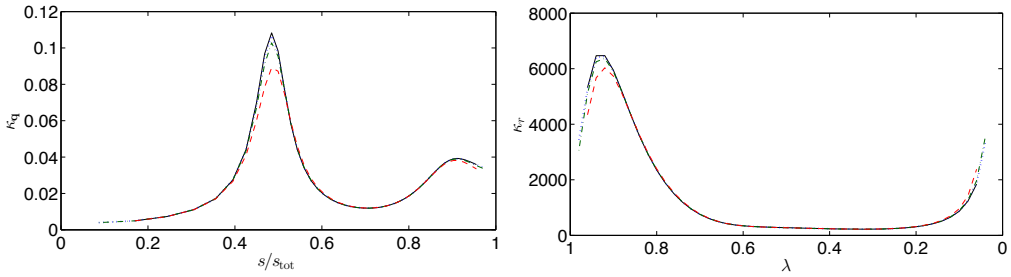


(d) Global homotopy

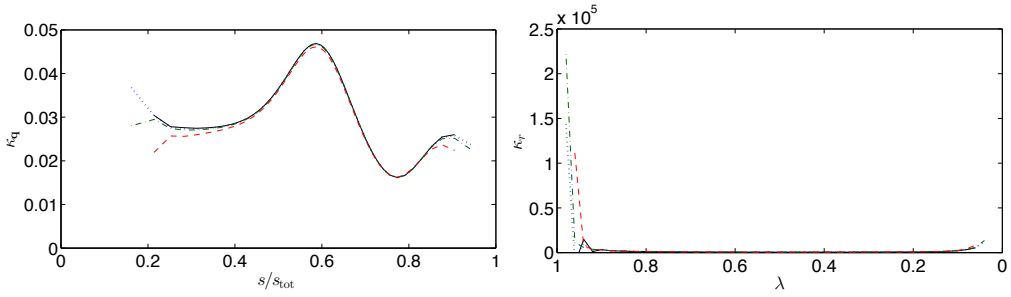
Figure A.22: Validation of the curvature estimation for inviscid flow over the ONERA M6 geometry on the fine mesh at Mach 0.4 and angle of attack of 3°



(a) Convex homotopy with dissipation operator with far-field boundary conditions

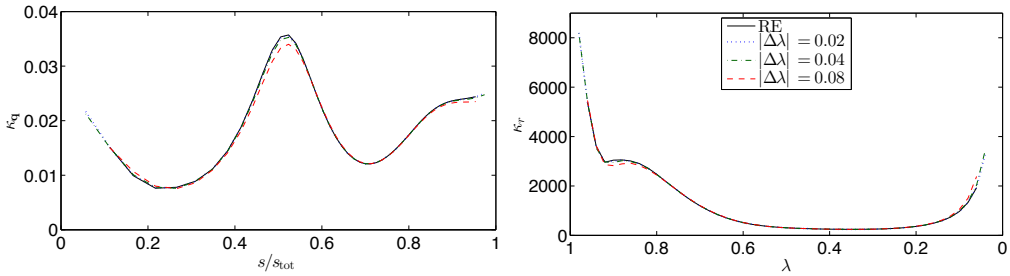


(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

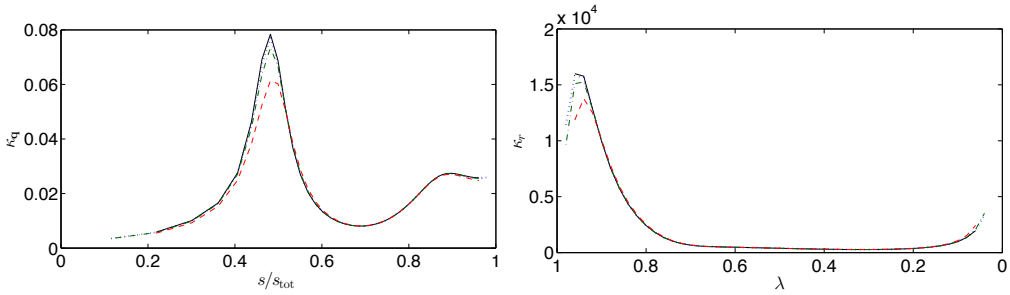


(c) Convex homotopy with diagonal operator

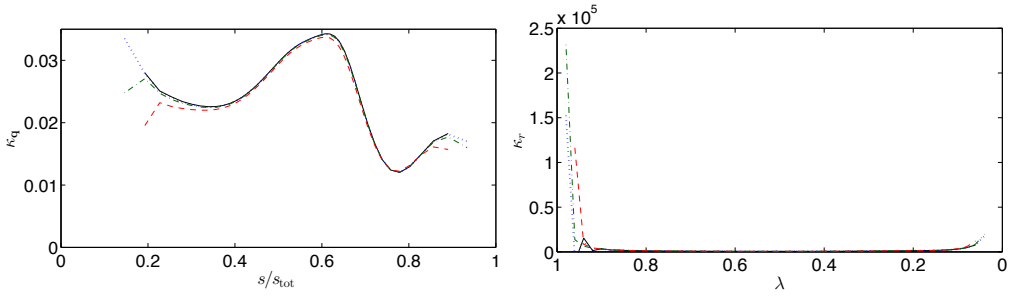
Figure A.23: Validation of the curvature estimation for the laminar ONERA M6 on the H-H topology mesh at Mach 0.3 and angle of attack of 1°



(a) Convex homotopy with dissipation operator with far-field boundary conditions

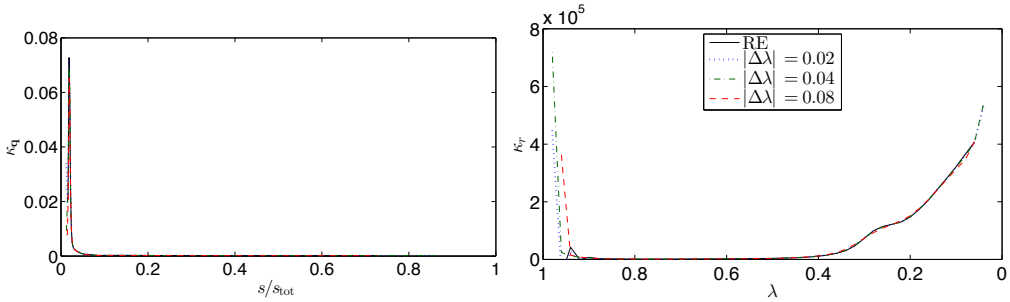


(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

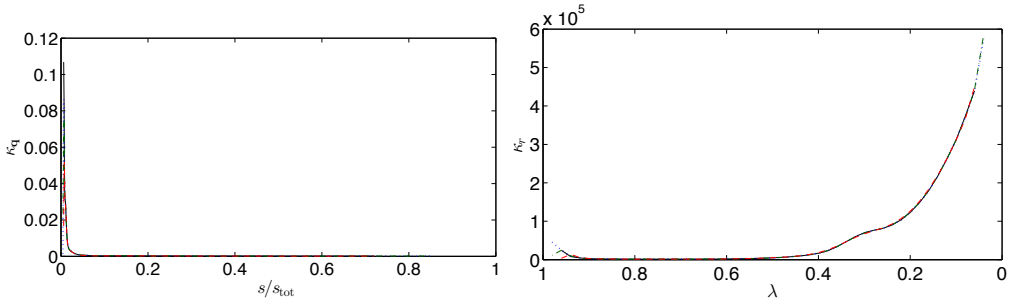


(c) Convex homotopy with diagonal operator

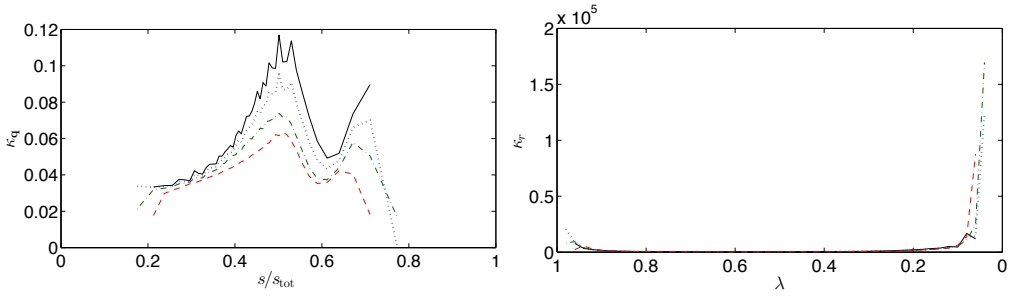
Figure A.24: Validation of the curvature estimation for the laminar ONERA M6 on the H-C topology mesh at Mach 0.3 and angle of attack of 1°



(a) Convex homotopy with dissipation operator with far-field boundary conditions

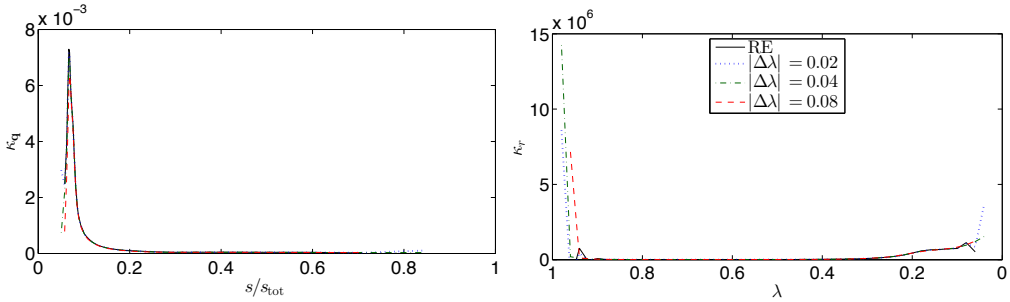


(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions

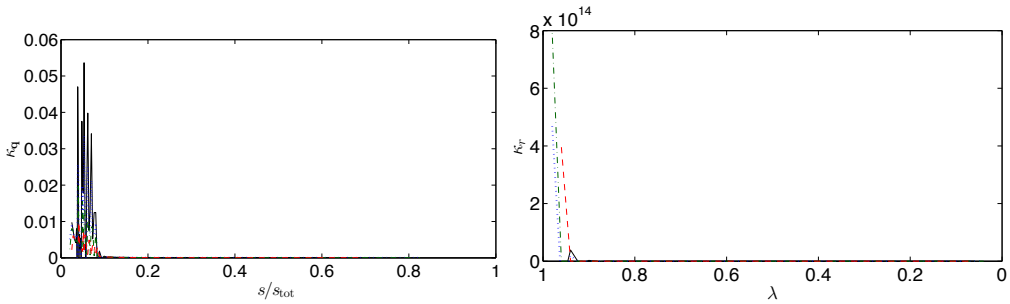


(c) Convex homotopy with diagonal operator

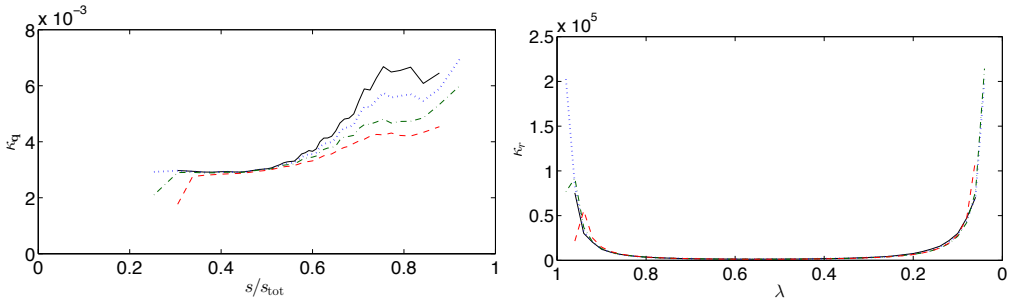
Figure A.25: Validation of the curvature estimation for turbulent flow over the NACA 0012 airfoil at Mach 0.4, Reynolds number of 4×10^6, and angle of attack of 1°



(a) Convex homotopy with dissipation operator with far-field boundary conditions



(b) Convex homotopy with dissipation operator with flow-imitative boundary conditions



(c) Convex homotopy with diagonal operator

Figure A.26: Validation of the curvature estimation for the ONERA M6 cases at Mach 0.4, Reynolds number of 10^6, and angle of attack of 3°