A Reduced-Order Model for Aerodynamic Shape Optimisation


Published in:
Aerospace Science and Technology

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2020 Elsevier Masson SAS. All rights reserved.
This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback
A Reduced-order Model for Gradient-based Aerodynamic Shape Optimisation

Weigang Yao\textsuperscript{a,1}, Simão Marques\textsuperscript{b,1,∗}, Trevor Robinson\textsuperscript{c,2}, Cecil Armstrong\textsuperscript{c,3}, Liang Sun\textsuperscript{d,4}

\textsuperscript{a}Faculty of computing, Engineering and Media, De Montfort University, Leicester, LE1,9BH, U.K.
\textsuperscript{b}Department of Mechanical Engineering Sciences, The University of Surrey, Guildford, England, GU2 7XH U.K.
\textsuperscript{c}School of Mechanical and Aerospace Engineering, Queen’s University Belfast, Belfast, Northern Ireland, BT9 5AH, U.K.
\textsuperscript{d}School of Computing, Engineering and Intelligent Systems, Ulster University, Northern Ireland, BT48 7JL, U.K.

Abstract

This work presents a reduced order model for gradient based aerodynamic shape optimization. The solution of the fluid Euler equations is converted to reduced Newton iterations by using the Least Squares Petrov-Galerkin projection. The reduced order basis are extracted by Proper Orthogonal Decomposition from snapshots based on the fluid state. The formulation distinguishes itself by obtaining the snapshots for all design parameters by solving a linear system of equations. Similarly, the reduced gradient formulation is derived by projecting the full-order model state onto the subspace spanned by the reduced basis. Auto-differentiation is used to evaluate the reduced Jacobian without forming the full fluid Jacobian explicitly during the reduced Newton iterations. Throughout the optimisation trajectory, the residual of the reduced Newton iterations is used as an indicator to update the snapshots and enrich the reduced order basis. The resulting multi-fidelity...
optimisation problem is managed by a trust-region algorithm. The ROM is demonstrated for a subsonic inverse design problem and for an aerofoil drag minimization problem in the transonic regime. The results suggest that the proposed algorithm is capable of aerodynamic shape optimization while reducing the number of full-order model queries and time to solution with respect to an adjoint gradient based optimisation framework.

*Keywords:* ROM, CFD, Aerodynamics, Shape Optimisation, Gradient-based Optimisation, Trust-Region, Multi-fidelity
Nomenclature

Latin Symbols

\( C_l \) = lift coefficient
\( C_d \) = drag coefficient
\( C_E \) = equality constraint
\( C_I \) = inequality constraint
\( C_m \) = pitching moment coefficient
\( D \) = diagonal matrix containing singular values
\( F \) = objective function
\( J \) = fluid Jacobian matrix
\( lb \) = design parameter lower bound values
\( M_\infty \) = free-stream Mach number
\( MUSCL \) = Monotonic Upwind Scheme for Conservation Laws
\( N \) = number of degrees of freedom
\( n_p \) = number of design parameters
\( n_r \) = number of basis
\( p \) = static pressure
\( \hat{p} \) = non-dimensional static pressure, \( \hat{p} = \frac{p}{q_\infty} \)
\( p \) = line search vector for Newton method
\( q \) = dynamic pressure
\( R \) = vector of fluid equations residuals
\( S \) = surface mesh points
\( s \) = optimisation step
\( t \) = time
\( U \) = left singular vectors
$ub$ = design parameter upper bound values

$V$ = right singular vectors

$V_n$ = design velocity

$w$ = vector of fluid and structural unknowns

$w_r$ = reduced vector of fluid unknowns

$X$ = matrix of snapshots

**Greek Symbols**

$\alpha$ = step length for Newton method, angle of attack

$\Delta$ = trust region radius

$\delta$ = small perturbation

$\varepsilon$ = residual threshold

$\varepsilon_\mu$ = trust-region termination threshold

$\varepsilon_{\nabla}$ = trust-region gradient threshold

$\zeta$ = constant for Carter condition

$\Lambda$ = vector of adjoint unknowns

$\lambda_i$ = singular value

$\eta_1, \eta_2$ = trust-region effectiveness thresholds

$\rho$ = trust-region effectiveness

$\Phi$ = vector subspace

$\mu$ = design parameters

$\Psi$ = reduced Jacobian matrix
1. Introduction

Numerical shape optimization usually requires the solution of parametric Partial Differential Equations (PDEs). This is a challenging process due to the high computational cost associated with having to interrogate a large and complex model multiple times. The number of evaluations of the full-order model (FOM) depends on the number of parameters, the number of constraints and objectives of the problem. For shape optimization problems this usually means several more parameters than functional constraints. Even for problems with a moderate number of parameters, the need for multiple model evaluations quickly makes it less attractive to deploy evolutionary methods or alternative global optimization algorithms [34]. Gradient-based optimization approaches, on the other hand, require the calculation of parametric sensitivity and respective function gradients. The cost of evaluating gradients, i.e. the scaling of the gradient evaluation with the number of parameters and associated number of model evaluations, can be mitigated by employing adjoint methods [30, 16, 17, 12, 28, 13]. This, however, introduces another PDE system with the same number of degrees of freedom as the primal PDE problem, equally or more challenging to solve.

Reduced-order modelling remains a popular topic in many engineering disciplines as a means to accelerate otherwise impractical or intractable simulations. Broadly speaking, the term reduced-order model (ROM) is a relative term that presupposes the existence of a FOM, of which some output is to be replicated at a reduced expense. More pertinent to this work, are the class of ROMs that maintain a close link to the physics described by the FOM but are computationally more efficient. This typically involves exam-
ining the governing equations and performing some type of model reduction, suitable for the physics of the problem. A particular model reduction technique assumes that the parametric behaviour of the FOM can be captured by a small number of modes or basis, typically obtained by methods such as Proper Orthogonal Decomposition (POD) [33], Balance POD [35, 32], Proper Generalized Decomposition [4]; excellent overviews of the range of applications and model reduction methods can be found in [22, 29, 20, 6, 5, 37]. In this work the resultant set of modes or basis is referred to as reduced order basis (ROB).

Within the context of model reduction, shape optimization involving nonlinear fluid equations poses significant challenges to model reduction due to the large number of parameters, more or less exacerbated by the degree of nonlinearity exhibited by the flow problem. Therefore, it is difficult to build a static and global basis that effectively covers the design space. Relevant efforts to this study include the approach for inverse aerofoil design of LeGresley and Alonso [21], where a POD ROM based on a Hicks-Henne surface parameterization was constructed to approximate the gradient by finite-differencing. Manzoni et al. successfully solve the Stokes equations for shape optimization of coronary arteries parameterized using a mapped Free-Form Deformation (FFD) technique and a reduced basis method [23]; the reduced basis are built from samples of the design space, obtained during an offline phase that incurs the majority of the computational cost. This technique was later expanded and applied to minimize vorticity by solving the same equations [27]. The elliptical nature of the PDE system enabled the authors to compute error bounds for the ROM. The availability of error
bounds for the surrogate model allowed Yue and Meerbergen [38] to prove that a trust-region method converges to the FOM optimum relying only on surrogates. This idea is further developed in [31], that proposed a trust region approach using reduced basis to build a ROM with a posteriori error bounds for elliptical and parabolic PDEs. Here, instead of building a global set of basis, the trust region method triggers updates to the basis when the error deteriorates, minimizing calls to the FOM.

For problems exhibiting stronger nonlinearities, building a global ROB, deducing error bounds for the ROM becomes increasingly challenging or is not feasible. A possible solution to this problem is to implement a zonal approach where the FOM is used to solve for the region of the flow subject to strong nonlinearities and a ROM reduces the overall cost by solving the remainder of the domain [15]. In the absence of strict error bounds, Zahr and Farhat assume a monotonic relationship between residual norm of the ROM and its error and proposed a nonlinear trust-region optimisation method that updates the ROB along the optimization trajectory, when the residual fails to reach a required threshold [39]. The authors exploit a Least-Squares Petrov-Galerkin projection [18] to reduce the state equations and respective sensitivities, hence, each snapshot involved concatenating samples of the fluid state variables and sensitivities with respect to the design variables.

To further reduce the number of snapshots required to build accurate ROMs, this work proposes a new type of ROM for gradient based aerodynamic shape optimization problems, centred on ROB built with samples from solving a linear system. As in reference [39], the residual is used to trigger the enrichment of the ROB within a trust-region framework, however, in this
work only sensitivity information is collected as part of each snapshot used to build both the fluid state and compute the gradient of the design variables. Further efficiency is gained by employing an auto-differentiation procedure to compute the reduced Jacobian. A subsonic inverse design and a transonic flow drag minimization problems are adopted to demonstrate the capability of the ROM based optimization methodology for problems exhibiting weak and strong nonlinearities. The efficiency of the proposed method is benchmarked against the FOM adjoint based optimisation strategy.

The paper is organized as follows: section 2 introduces the reference optimisation framework using the FOM for analysis and respective adjoint based gradient computation; section 3 derives the ROM, describes the ROBs updating procedure and the ROM gradient evaluation required for the optimisation; section 4 introduces the trust-region formalism used to solve the optimisation problems employing the ROM; this is followed by two sets of results and respective analysis; finally, the paper finishes with a conclusion section.

2. Shape Optimization Using a Full-Order Model

The FOM used to obtain the flow solutions in this work solves the compressible fluid Euler equations. The equations are discretized using a cell-centred finite-volume scheme over block-structured conforming meshes, employing a Roe flux function, together with MUSCL interpolation, and the van Albada limiter is used to obtain second-order accuracy. The nonlinear system of algebraic equations is marched forward in time by an explicit four-stage Runge-Kutta method [36].
The flow solver provides the optimisation objective and constraints as a function of a given geometry. The geometry itself, is parameterised in terms of design parameters that allow defining the shape of interest. Hence, the optimization problem can be formulated as:

\[
\begin{align*}
\text{minimize}_{\mu \in \mathcal{D}} & \quad F(w(\mu), \mu) \\
\text{subject to} & \quad R(w(\mu), \mu) = 0, \\
& \quad C_E(w(\mu), \mu) = 0, \\
& \quad C_I(w(\mu), \mu) \leq 0,
\end{align*}
\]

where \( F \) is the objective function, \( w \) and \( \mu \) are fluid state and design parameters, respectively, \( \mathcal{D} \subset \mathbb{R}^n \) represents the parameter space, \( R \) is the residual of the Euler equations and \( C_E, C_I \) are equality and inequality constraints, respectively. To solve for the fluid state variables, the fluid residual \( R(w(\mu), \mu) \) is driven to zero.

For gradient-based optimisation, the total derivative can be used to express how the objective function changes with respect \( \mu \):

\[
dF = \frac{\partial F}{\partial \mu} + \frac{\partial F}{\partial w} dw.
\]

Note that while computing Eq. 2, \( R(w(\mu), \mu) = 0 \) must still be satisfied, therefore we can use

\[
dR = \frac{\partial R}{\partial \mu} + \frac{\partial R}{\partial w} dw = 0.
\]

The cost of evaluating \( \frac{dw}{d\mu} \) is not trivial as it requires the solution of the
following linear system

\[
\frac{d\mathbf{w}}{d\mu} = - \left[ \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \right]^{-1} \frac{\partial \mathbf{R}}{\partial \mu},
\]  

(4)

where the term \( \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \in \mathbb{R}^{N \times N} \), is the fluid Jacobian \( \mathbf{J} \), with \( N \) representing the degrees of freedom of the fluid state. Substituting Eq. 4 into Eq. 2, results in

\[
\frac{d\mathbf{F}}{d\mu} = \frac{\partial \mathbf{F}}{\partial \mu} - \frac{\partial \mathbf{F}}{\partial \mathbf{w}} \left[ \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \right]^{-1} \frac{\partial \mathbf{R}}{\partial \mu}.
\]  

(5)

The evaluation of the derivatives in Eq. 5 can be performed by finite-differencing, however this quickly becomes impractical even for a small number of design variables. Alternatively, it is possible to employ the so called direct or adjoint methods. This requires the solution of an appropriate linear system, see for example Hwang and Martins for further details [24]. The direct method requires the solution of Eq. 4, the cost of solving this linear system is proportional to the number of design variables. The adjoint method involves re-writing Eq. 5 as

\[
\frac{d\mathbf{F}}{d\mu} = \frac{\partial \mathbf{F}}{\partial \mu} + \Lambda^T \frac{\partial \mathbf{R}}{\partial \mu},
\]  

(6)

where \( \Lambda^T \) is the adjoint vector, obtained by solving the following linear (adjoint) system,

\[
\left[ \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \right]^T \Lambda = - \left[ \frac{\partial \mathbf{F}}{\partial \mathbf{w}} \right]^T.
\]  

(7)

The Jacobian matrix \( \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \) and right hand side vector \( \frac{\partial \mathbf{F}}{\partial \mathbf{w}} \) in Eq. 7 are obtained through auto-differentiation of the discretized Euler equations using the Tapenade library [14]. The Jacobian is typically a large, sparse matrix where only the non-zeros entries are computed. For small problems, matrix factorization can be used to solve Eq. 7, which is the case in this work;
alternatively, iterative solvers such as GMRES can be used for larger meshes, whereby only matrix-vector products are necessary to solve the linear system.

The adjoint system scales with the number of functions in Eq. 1, therefore, is the preferred choice for aerodynamic shape optimisation problems, where the number of parameters, $n_p$, usually outnumbers the number of objective and constraint functions. In this work, the optimisation results based solely on the FOM are used as reference and are solved using the Sequential Quadratic Programming (SQP) algorithm from MatLab’s *fmincon* function [26], with the gradient built using the solution of the adjoint system represented in Eq. 7.

3. Shape Optimization Using a Reduced-Order Model

The aim of this work is to reduce the cost of solving Eq. 1, by employing an approximate model, i.e. the ROM, in lieu of the FOM, at minimum loss of fidelity. To minimize cost, the ROM is built as the optimisation progresses, i.e. the ROM is only developed along the optimisation trajectory. This work aims to minimize or mitigate the dependence of the number of model queries on the number of design variables required to either build the ROM or compute gradients. The result is a multi-fidelity optimisation formulation, employing a trust-region strategy to manage the high and low fidelity model solves.

3.1. FOM Reduction

The central idea of projection-based ROMs is to project the full state vector $\mathbf{w}$ onto a subspace $\Phi \in \mathbb{R}^{N \times n_r}$ spanned by the ROB, where $N \gg n_r$,
i.e.:

\[ \mathbf{w} \approx \tilde{\mathbf{w}} + \Phi \mathbf{w}_r, \]  

(8)

where \( \tilde{\mathbf{w}} \) is the operation reference state and \( \mathbf{w}_r \) is the reduced state vector. By Substituting Eq. 8 into the fluid residual \( \mathbf{R} \), the FOM is converted to the following optimization problem,

\[
\text{minimize} \quad L_2 = \frac{1}{2} \| \mathbf{R}(\tilde{\mathbf{w}} + \Phi \mathbf{w}_r, \mu) \|_2^2.
\]  

(9)

This type of ROM is equivalent to the minimum residual approach, meaning that if the ROB is enriched, the solution error in the \( \mathbf{J}^T \mathbf{J} - \text{norm} \) is non-increasing [8]. The first order optimality condition gives

\[
\frac{dL_2}{d\mathbf{w}_r} = \Psi^T \mathbf{R}(\tilde{\mathbf{w}} + \Phi \mathbf{w}_r, \mu) = 0,
\]  

(10)

where \( \Psi = \mathbf{J} \Phi \in \mathbb{R}^{N \times n_r} \) is the reduced Jacobian matrix. Following [18], the Newton method is employed to solve Eq. 10, referred to as the reduced Newton iterations:

\[
\Phi^T \mathbf{J}_k^T \mathbf{J}_k \Phi \mathbf{p}_k = -\Phi^T \mathbf{J}_k^T \mathbf{R}_k 
\]  

(11)

\[
\mathbf{w}_r^{k+1} = \mathbf{w}_r^k + \alpha_k \mathbf{p}_k. 
\]  

(12)

In the above reduced Newton method, also known as Least-Squares Petrov-Galerkin projection formulation, \( \alpha_k \) is the step length and \( \mathbf{p}_k \) is the line search direction vector. The reduced Jacobian \( \mathbf{J} \Phi \) needs to be evaluated at each iteration, which is the major computational cost and requires access to the solver functions. As mentioned above, the Jacobian matrix is typically a large and sparse matrix, scalable with \( \mathbf{w} \) degrees of freedom. Evaluating and storing \( \mathbf{J} \) explicitly quickly becomes impractical for large scale problems, By
realizing that

\[ \mathbf{J} \Phi = \delta \mathbf{R}(\bar{\mathbf{w}}) \Phi \approx \frac{1}{\delta} [\mathbf{R}(\bar{\mathbf{w}} + \delta \Phi, \mu) - \mathbf{R}(\bar{\mathbf{w}}, \mu)] \] (13)

it is apparent that \( \mathbf{J} \Phi \) can be approximated by finite-differences, however this would require the user to set the perturbation parameter, \( \delta \). This can be avoided by using auto-differentiation. Recall that \( \Phi \in \mathbb{R}^{N \times n_r} \), therefore forming \( \mathbf{J} \Phi \) requires \( N \times n_r \) flux function calls. The value \( (N \times n_r) \) is, in general, much smaller than the number of non-zero entries in \( \mathbf{J} \), i.e. the number of flux function evaluations required to explicitly compute \( \mathbf{J} \), which suggests auto-differentiation methods become more advantageous for larger problems, therefore it is the preferred approach in this work.

### 3.2. Construction and Enrichment of Reduced Order Basis

It is common practice to extract ROB from a matrix of snapshots or database \( \mathbf{X} \) by POD [33]. The key is to build sufficient data in the range of interest. In the present work, only snapshots of \( \frac{\partial \mathbf{w}}{\partial \mu} \) samples are considered, assuming the nonlinearity is weak or the state does not deviate far from the operation state \( \bar{\mathbf{w}} \). The snapshots are computed by solving the linear system of equations given by Eq. 4, that is,

\[ \mathbf{X} = \frac{d \mathbf{w}}{d \mu} = - \left[ \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \right]^{-1} \left[ \frac{\partial \mathbf{R}}{\partial \mu} \right]. \] (14)

In Eq. 14, \( \frac{\partial \mathbf{R}}{\partial \mathbf{w}} \) is computed from the fluid system and

\[ \frac{\partial \mathbf{R}}{\partial \mu} = \frac{\partial \mathbf{R}}{\partial \mathbf{S}} \frac{\partial \mathbf{S}}{\partial \mu} \] (15)

where \( \mathbf{S} \) represents the surface mesh and \( \frac{\partial \mathbf{S}}{\partial \mu} \) is obtained from the parameterization of the problem, e.g. Class-Shape Transformation (CST) [19], FFD [24]
or third party CAD systems [2]; in this work the mesh is deformed following the transfinite interpolation method and the term \( \frac{\partial R}{\partial S} \) which represents the flow residual sensitivity to mesh changes is obtained by auto-differentiation.

Once the database is obtained, the following procedure is adopted to build the subspace \( \Phi \):

\[
X^T X = U D V^T, \quad \text{(16)}
\]

\[
\Phi = X V D^{-\frac{1}{2}}. \quad \text{(17)}
\]

Where \( D = \text{diag}[\lambda_1, \lambda_2, ..., \lambda_{np}] \) corresponds to the singular values, and \( U, V \) are the left and right singular vectors, respectively. The cost of applying singular value decomposition (SVD) to \( X^T X \), which is an \( n_r \times n_r \) matrix, is trivial.

It is difficult, if possible at all, to construct a static global subspace \( \Phi \) for a nonlinear system, therefore, it is necessary to update \( \Phi \) along the optimization trajectory. The \( L_2 \) norm of the fluid residual in the Newton’s iteration is used as an indicator to update \( \Phi \). If the \( L_2 \) norm remains larger than the user defined threshold \( \varepsilon \), new snapshots are generated by Eq. 14 at the current state and appended to the previous snapshot matrix, and used to enrich \( \Phi \).

### 3.3. Gradient Evaluation

By substituting Eq. 8 into the FOM gradient Eq. 4 yields the reduced gradient approximation:

\[
\frac{d w}{d \mu} \approx -\Phi \left[ J \Phi \right] + \frac{d R}{d \mu}, \quad \text{(18)}
\]
Where "+" denotes Moore-Penrose pseudoinverse. Substituting Eq. 18 into Eq. 2 and obtain

\[
\frac{dF}{d\mu} \approx \frac{\partial F}{\partial \mu} - \frac{\partial F}{\partial w} \Phi [J\Phi]^+ \frac{dR}{d\mu}.
\] (19)

Eq. 19 can be expressed as the following by introducing the design velocity \( V_n = \frac{dS}{d\mu} \), which is the surface mesh derivative with respect to the design parameter \( \mu \) [2],

\[
\frac{dF}{d\mu} \approx [\frac{\partial F}{\partial S} - \frac{\partial F}{\partial w} \Phi [J\Phi]^+ \frac{\partial R}{\partial S}] V_n. 
\] (20)

The objective function sensitivity to the mesh changes, \( \frac{\partial F}{\partial S} \), is obtained from the auto-differentiation of the solver.

Compared with the full order adjoint system in Eq. 7, which requires solving an \( N \times N \) linear system, the reduced system only needs to solve an \( N \times n_r \) linear system, which provides the speed up in the gradient computation. However, it is necessary to point out that the \( \Phi \) updates scale with the number of design variables due to the need to solve Eq. 14, hence this method remains competitive if the number of \( \Phi \) updates remains significantly lower than the FOM calls required by the adjoint based method.

4. Trust-Region Framework

As the name indicates, the trust-region method aims to establish a subdomain where the low-fidelity model provides an adequate representation of the FOM [10]. At each major iteration \( k \), an optimisation subproblem is
defined on the trust-region centred at \( \mu_k \) and radius \( \Delta_k \):

\[
\begin{align*}
\minimize_{s \in \mathcal{B}_k} & \quad \hat{\mathcal{F}}(\mathbf{w}(\mu_k), \mu) \\
\text{subject to} & \quad \hat{\mathcal{C}}_E(\mathbf{w}(\mu_k), \mu_k + s) = 0, \\
& \quad \hat{\mathcal{C}}_I(\mathbf{w}(\mu_k), \mu_k + s) \leq 0,
\end{align*}
\]

\( \text{lb} \leq (\mu_k + s) \leq \text{ub}, \)

\[ ||s||_\infty \leq \Delta_k, \]

where the “\(^\hat{\ }\)” symbol indicates quantities computed using the ROM (using Eq. 8 for the functionals and Eq. 20 for the gradients); \( s \) is the optimisation step size and \( \mathcal{B}_k = \{ \mu \in \mathbb{R}^{np} : ||\mu - \mu_k|| \leq \Delta_k \} \).

Trust region methods are provably convergent for constrained optimisation problems to an optimum of the FOM, provided both models satisfy a number of conditions, including that the low-fidelity model is corrected to be at least first-order consistent with the FOM [3, 7, 1], i.e.:

\[
\begin{align*}
\mathcal{F}(\mu_k) &= \hat{\mathcal{F}}(\mu_k), \mathcal{C}_E(\mu_k) = \hat{\mathcal{C}}_E(\mu_k), \mathcal{C}_I(\mu_k) = \hat{\mathcal{C}}_I(\mu_k) \\
\nabla \mathcal{F}_k &= \nabla \hat{\mathcal{F}}_k, \nabla \mathcal{C}_E(\mu_k) = \nabla \hat{\mathcal{C}}_E(\mu_k), \nabla \mathcal{C}_I(\mu_k) = \nabla \hat{\mathcal{C}}_I(\mu_k).
\end{align*}
\]

For unconstrained optimisation, the first-order consistency requirement can be relaxed and a suitable approximation of the gradient at the centre of the trust region will suffice. This approximation is usually based on the Carter condition,

\[
\frac{||\nabla \mathcal{F}_k - \nabla \hat{\mathcal{F}}_k||}{||\nabla \hat{\mathcal{F}}_k||} \leq \zeta, \quad \forall k,
\]

with the constant \( \zeta < 1, \) [9, 11].

To guarantee convergence to a stationary point of the FOM, the following are also required [10]: a) \( \hat{\mathcal{F}} \) is locally Lipschitz continuous and regular with
respect to \( s \) for all \( \mu \) and continuous in \( \mu \) for all \( s \); b) the set of problem parameters is closed and bounded; c) the sufficient decrease condition requires the step to satisfy the fraction of Cauchy decrease (FCD). In addition, the second derivatives of the ROM at \( \mu_k \) remain bounded within the trust region domain for all \( k \) \cite{1}. The FCD condition was derived for the classical trust-region method, where the FOM is approximated by a quadratic Taylor’s series expansion. With POD based approaches, this is no longer feasible, instead the low-fidelity function step is determined by solving Eq. 21 using MatLab’s \texttt{fmincon} SQP algorithm, which maximizes the decrease in the objective function of the trust-region subproblem.

The effectiveness of the trust-region step is evaluated by the ratio of the actual improvement over the improvement predicted by the ROM, \( \rho \):

\[
\rho_k = \frac{\mathcal{F}(\mu_k) - \mathcal{F}(\mu_k + s)}{\hat{\mathcal{F}}(\mu_k) - \hat{\mathcal{F}}(\mu_k + s)}. \quad (25)
\]

For values of \( \rho < \eta_1 \) the step is rejected and the trust-region radius is reduced; if \( \eta_1 < \rho < \eta_2 \), the trust-region size is maintained and increased when \( \rho > \eta_2 \). In this work, the trust-region is set up with \( \eta_1 = 0.5 \) and \( \eta_2 = 0.9 \); the trust-region size, \( \Delta_{k+1} \), is then reduced by a factor of 0.5, maintained or increased by a factor 1.25.

The optimisation terminates when the change in the design variables is less than the termination threshold, \( \varepsilon_\mu \), or the trust-region size drops below a minimum \( \Delta_k < \Delta_{\text{min}} \). This is complemented by enforcing the assessment of the ratio \( \rho \) when the ROM gradient norm is below the threshold \( \varepsilon_{\nabla} \). Hence, if the ROM is not accurate when the ROM optimiser is in effect converged, the trust-region radius is reduced, which means that eventually the linear approximation will be valid and the FOM gradient norm will also drop below
the $\varepsilon$. This is particularly relevant for highly nonlinear problems, where the ROM accuracy is more volatile. Algorithm 1 summarises the complete process.

5. Results

5.1. Subsonic Flow Inverse Design

A subsonic inverse design test case used in [39] was chosen to assess the current ROM, the problem is formulated as:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| p(w, \mu) - p_{RAE2822} \|_2^2 \\
\text{subject to} & \quad R(w(\mu), \mu) = 0
\end{align*}
\]

where $p$ is the pressure distribution. Equation 26 aims to modify the aerofoil shape to match the target pressure distribution $p_{RAE2822}$ produced by an inviscid analysis of the $RAE2822$ aerofoil, hence the optimum shape should correspond to the $RAE2822$ aerofoil. The flow is described by the fluid Euler equations, the free-stream conditions are $(M_\infty, \alpha) = (0.5, 0^\circ)$. The $NACA0012$ aerofoil is used as the initial geometry. The CST method using 20 weights ($n_p = 20$), is used to parameterize the aerofoil by superimposing Bernstein polynomials, shown in Fig. 1, on the initial geometry definition. The CST formulation also provides the analytical definition of $V_n$. Figure 2-(a) shows the aerodynamic grid around the $NACA0012$ aerofoil and Fig. 2-(b) the respective pressure field. The O-type grid contains $121 \times 41$ points in the circumferential and radial directions respectively, resulting in the fluid state $w$ of dimension $N = 19844$. 

18
Algorithm 1 ROM for Shape Optimization

**Input:** Initial geometry, initial parameters values, ROM update threshold $\varepsilon$, ROM maximum order, $\Delta_0$

**Output:** Optimum shape, fluid state $w$

1: **Initialize:**
2: compute snapshot using Eq. 14, compute $\Phi$ with Eq. 17
3: **while** $\Delta_k > \Delta_{min}$ **do**
4: solve optimisation subproblem, Eq. 21
5: **if** Fluid residual $L_2 > \varepsilon$ or $|\nabla_{ROM}| < \varepsilon$ **then**
6: compute $\rho$
7: **if** $\rho < \eta_1$ **then**
8: reject step, update $\Delta_k(\eta_1)$
9: **else**
10: accept step, update $\Delta_k(\eta_1, \eta_2)$
11: $\mu_{k+1} \leftarrow \mu_k + s$
12: compute new snapshot using Eq. 14, enrich basis, update $\Phi$
13: compute $\nabla \hat{F}(\mu_{k+1})$ if corrections are required.
14: **end if**
15: **else**
16: accept step, increase $\Delta_k$
17: $\mu_{k+1} \leftarrow \mu_k + s$
18: **end if**
19: **if** $|\mu_{k+1} - \mu_k| < \varepsilon_\mu$ **then**
20: **stop**
21: **end if**
22: **end while**
Figure 1: Bernstein polynomials used to parameterise the geometry.

Figure 2: (a) aerodynamic grid of NACA0012 aerofoil; (b) non-dimensional pressure, $\hat{p}$, flow field at $(M_\infty, \alpha) = (0.5, 0)$. 
Table 1: Relative cost to compute the ROM components, normalized by the wall-clock time required to solve the steady state once.

<table>
<thead>
<tr>
<th>$\bar{w} + \Phi$</th>
<th>$J\Phi$</th>
<th>$(\frac{\partial p}{\partial w}, \frac{\partial R}{\partial S})$</th>
<th>$w_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.75</td>
<td>0.05</td>
<td>0.12</td>
<td>0.25–0.4</td>
</tr>
</tbody>
</table>

The grid size is normalized by the aerofoil chord length $c$. Following Eq. 19, the pressure gradient with respect to the design parameters is given by

$$\frac{dp}{d\mu} = -\frac{\partial p}{\partial \bar{w}} J^{-1} \frac{\partial R}{\partial S} V_n, \quad (27)$$

$$\approx -\frac{\partial p}{\partial \bar{w}} \Phi \left[ J\Phi \right]^+ \frac{\partial R}{\partial S} V_n. \quad (28)$$

The subspace $\Phi$ is constructed using POD on the snapshot matrix defined by Eq. 14. The process starts with 20 ROBs being retained for the first optimization iteration, and is limited to 40 ROBs for the remainder iterations. Table 1 shows the wall clock time required to build the ROM and reconstruct the fluid state. It is worth noting that the cost of building $\Phi$ requires computing the steady state, i.e. the reference state $\bar{w}$, extracting the fluid Jacobian and solving the $n_p$ linear system of Eq. 4. The cost of evaluating $w_r$ increases with the addition of basis to the ROM.

Figure 3 shows the gradient calculated for the initial conditions using finite differencing (FD) and the adjoint FOM (Eq. 6), together the with the ROM prediction obtained from Eq. 28. All methods are in excellent agreement with each other, indicating the Carter condition is respected at the centre of the trust region. The trust-region setup for this problem avoids the use of corrections and assumes the gradient approximation satisfies the Carter condition, results in Table 2 show the Carter condition is indeed met.
Table 2: Carter condition parameter, $\zeta$, and trust-region effectiveness, $\rho$.

<table>
<thead>
<tr>
<th>ROM update</th>
<th>$\zeta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1317</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>0.0648</td>
<td>1.0570</td>
</tr>
<tr>
<td>3</td>
<td>0.0686</td>
<td>0.7916</td>
</tr>
<tr>
<td>4</td>
<td>0.0262</td>
<td>1.0312</td>
</tr>
<tr>
<td>5</td>
<td>0.0307</td>
<td>1.0226</td>
</tr>
<tr>
<td>6</td>
<td>0.0386</td>
<td>0.9949</td>
</tr>
<tr>
<td>7</td>
<td>0.0304</td>
<td>0.9635</td>
</tr>
<tr>
<td>8</td>
<td>0.0447</td>
<td>0.9702</td>
</tr>
<tr>
<td>9</td>
<td>0.0280</td>
<td>1.0221</td>
</tr>
<tr>
<td>10</td>
<td>0.3216</td>
<td>0.9697</td>
</tr>
<tr>
<td>11</td>
<td>0.2381</td>
<td>0.9462</td>
</tr>
</tbody>
</table>

The threshold $\varepsilon$ determines when $\Phi$ is updated, a smaller value of $\varepsilon$ results in more frequent updates, hence more FOM calls, as shown by Fig. 4-(a). A ROM with $\varepsilon = 10^{-5}$ requires three updates. However, it is worth noting the threshold is case dependent and must be defined a priori by the user. In this case, the a value of $\varepsilon = 10^{-7}$ is used to complete the assessment of the ROM. The gradient norm predicted by the ROM at the centre of each trust-region is compared to the FOM gradient in Fig. 4-(b), indicating the level of agreement shown in Fig. 3 is maintained throughout the optimisation trajectory.

The optimisation trajectory and final geometry obtained are shown in Fig.
Figure 3: Pressure gradient with respect to the first weight comparison. The Finite-Differencing (FD) result is computed by perturbing the first weight with a step-size $\delta = 0.01$. The deformed shape is enlarged by 10 times for better visualization. The ROM used 20 basis.
Figure 4: (a) Convergence of the objective function for different residual thresholds, ε; (b) objective function gradient norm evaluated at the centre of each trust-region.

5 using the residual threshold of $\varepsilon = 10^{-7}$. Results show the ROM converging to the FOM optimum, with both methods requiring just over 100 iterations to reach the optimum. As shown in table. 3, the ROM requires 11 steady state evaluation and linear snapshots to be collected to match the target pressure, which compares favorably with respect to the FOM optimization. The cost of using the ROM to evaluate the objective function and gradient is not trivial, the ROM iterations require the evaluation of $\mathbf{J}\mathbf{\Phi}$, which dominates the computational effort of each iteration (about 95%), the remainder operations to compute $\mathbf{w}_r$ are two orders of magnitude faster using QR decomposition MatLab function $qr[25]$; in the end, a reduction to approximately 70% of the total time required by the FOM based optimisation was achieved, when using $\varepsilon = 10^{-7}$. The pressure fields obtained from both the final geometries are compared in Fig. 6, which further underlines the ROM accuracy. The
fluid system is basically linear or rather weakly nonlinear, this resulted in an optimisation trajectory without any rejected steps, despite the ever increasing trust-region radius.

5.2. Transonic Flow Drag minimization

The RAE2822 aerofoil constrained drag minimization at \((M_\infty, \alpha) = (0.73, 2^\circ)\) problem is adopted to evaluate the proposed ROM for aerodynamic shape optimization in the transonic regime. The constrained optimisation problem is defined as follows:

\[
\begin{align*}
\text{minimize} & \quad C_d \\
\text{subject to} & \quad \mathbf{R}(\mathbf{w}(\mu), \mu) = 0 \\
& \quad (C_l, C_m) = (C_l, C_m)_{\text{RAE2822}}
\end{align*}
\]
Figure 6: Pressure field comparison between FOM and ROM ($\varepsilon = 10^{-7}$) final aerofoils. The solid and dashed lines represent ROM and FOM, respectively. The ROM result is produced by the FOM using with the aerofoil shape from the ROM optimization.

Table 3: Performance and resource usage - comparison of FOM and ROM based inverse design problem.

<table>
<thead>
<tr>
<th></th>
<th>FOM Optimisation</th>
<th>ROM Optimisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n. evals.</td>
<td>wall clock [s]</td>
</tr>
<tr>
<td>FOM steady state</td>
<td>105</td>
<td>1039</td>
</tr>
<tr>
<td>FOM adjoint state</td>
<td>105</td>
<td>1040</td>
</tr>
<tr>
<td>Update $\Phi$:</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ROM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total Wall Clock:</td>
<td>2079</td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{2}|p(w, \mu) - p_{RAE2822}|_2^2$:</td>
<td>$3.56 \times 10^{-4}$</td>
<td>$3.53 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Figure 7: (a) aerodynamic grid of RAE2822 aerofoil and (b) pressure flow field at $(M_\infty, \alpha) = (0.73, 2^\circ)$.

As in the previous case, a CST parameterization with 20 weights ($n_p = 20$) is used to control the shape. As shown in Fig. 7-(a), the solutions are obtained on an O-type grid with $161 \times 41$ points in the circumferential and radial directions, respectively. The resultant fluid state $w$ has a dimension of $N = 26404$. The initial pressure flow field is shown in Fig. 7-(b), which exhibits a shock just aft of the mid chord on the upper surface.

The gradients of the objective function, lift and pitching moment constraints are compared in Fig. 8, showing excellent agreement between the different methods available to compute the gradients. For the constrained optimisation problem, it was advantageous to include the corrections described by Eq. 22. Although numerical experiments show the problem reaches consistent converged solutions without corrections, the addition of these sped up convergence and yielded further reductions of the objective function.
Figure 8: Constrained optimisation gradient comparison - the finite-difference result is computed by perturbing the design parameters with a step-size $\delta = 0.001$. The ROM used 20 basis.
The presence of a strong nonlinearity in this problem required the use of a smaller residual threshold, $\varepsilon = 10^{-8}$, to produce meaningful solutions; however, the total number of $\Phi$ updates remains similar to the previous problem. The convergence of the ROM based optimisation is compared to the FOM using the adjoint method to compute the gradient in Fig. 10-(a). As in the previous test, the ROM is able to find a minimum close to the FOM based optimisation solution. Reducing the number of ROBs used to build the ROM deteriorates the convergence rate as illustrated in Fig. 10-(b). Nevertheless, the final solution is identical to those retaining 40 or 80 ROBs. Increasing the number of ROBs beyond 40 appears to have a limited impact on the solution.

The number of FOM calls for both optimisation strategies plotted in Fig. 10-(a) are shown in Table. 4, together with wall clock times. Using the matrix factorisation method to solve the adjoint systems greatly reduces the solution time required to obtain the gradients for the FOM optimisation with respect to the flow solution. For the ROM based optimisation, and as before, the number of FOM calls is reduced by almost a quarter. Despite the number of FOM calls reducing significantly for the ROM problem, the cost of evaluating the ROM using 40 basis leads only to a marginal reduction in wall-clock times for both strategies. However, without impacting the quality of final solution, it is possible to solve this problem retaining 20 ROBs, which yields more significant savings. To understand the impact of the number of ROB and design variables on the performance of this ROM based optimisation, the computational time to evaluate a functional with the ROM and updating $\Phi$ is compared in Fig. 9. For a given number of design
parameters, increasing number of ROBs has a dramatic effect on the effort required by the ROM - Fig. 9-(a), this stems mainly from the need to evaluate the reduced Jacobian $J\phi$, which is an operation $\propto (N \times n_r)$; on the other hand, the cost of evaluating the ROM is independent of the number of design parameters, the cost of updating $\Phi$, for a fixed number of ROBs, increases linearly with the number of design parameters, as shown in Fig. 9-(b). It is worth pointing out that the number of basis required to solve a problem adequately is dependent on the number of design parameters and as shown by the results, increasing the number of basis beyond 40 will not result in significant improvements over the FOM approach. As noted in [39], this excessive cost in evaluating functionals using the ROM can be mitigated by employing hyper-reduction techniques. Figure 11 shows the solutions corresponding to the minima found by the FOM and ROM optimisations. The results show that even for problems exhibiting nonlinearities such as shocks, the ROM based strategy is able to produce similar results to those obtained with conventional FOM optimisation strategies.

6. Conclusion

A ROM is developed for gradient based aerodynamic shape optimization with reduced Newton’s iterations. Auto-Differentiation is used to evaluate the reduced Jacobian without forming the full fluid Jacobian explicitly during the reduced Newton’s iterations. A sampling procedure based on the solution of linear system of equations, is adopted to construct the subspace. The procedure only requires solving linear systems with the number of design parameters without resorting to sample for new fluid states for each
Figure 9: computational effort for the ROM steady state solution and updating the $\Phi$: (a) as a function of the number of ROBs retained, for 20 design parameters; (b) as a function of the number of design parameters, whilst retaining 40 ROBs. Time is normalized with respect to cost one steady state solution.
Figure 10: (a) ROM and FOM based optimisation convergence history; (b) impact of number of ROBs on optimisation convergence.

Figure 11: Pressure field comparison between FOM and ROM ($\varepsilon = 10^{-8}$). The ROM result is obtained by FOM steady state solution with the aerofoil shape from the ROM optimization. The solid and dashed lines represent ROM and FOM, respectively.
Table 4: Performance and resource usage - comparison of FOM and ROM based transonic drag minimization.

<table>
<thead>
<tr>
<th></th>
<th>FOM</th>
<th>ROM 20 ROB</th>
<th>ROM 40 ROB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n. evals.</td>
<td>time[s]</td>
<td>n. evals.</td>
</tr>
<tr>
<td>FOM steady state</td>
<td>74</td>
<td>1952</td>
<td>14</td>
</tr>
<tr>
<td>FOM adjoint state</td>
<td>18</td>
<td>242</td>
<td>-</td>
</tr>
<tr>
<td>Update Φ:</td>
<td>-</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>ROM</td>
<td>-</td>
<td>-</td>
<td>142</td>
</tr>
<tr>
<td>Total wall clock time:</td>
<td>2194</td>
<td>1554</td>
<td>1825</td>
</tr>
<tr>
<td>$C_d$ reduction:</td>
<td>48.10%</td>
<td>48.17%</td>
<td>48.33%</td>
</tr>
</tbody>
</table>
design variable, which quickly becomes prohibitive for large number of design parameters. It was demonstrated that the proposed ROM is suitable for subsonic inverse design and transonic flow drag minimization problems. The inverse design optimization problem can be solved with fewer basis updates than the transonic case, as the latter represents a strongly nonlinear system. Even for the more demanding test case, it was possible to reduce the number of FOM analysis by a factor of four, however this was not translated into significant time savings.

Acknowledgments

The funding provided for this research from the Engineering and Physical Sciences Research Council (Grant No. EP/P025692/1) is gratefully acknowledged.


