The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-–9 June 2022, Oslo, Norway

ACCELERATED NONLINEAR-PDE-CONSTRAINED OPTIMIZATION BY REDUCED ORDER MODELLING

BENJAMIN F. GIBSON[1] AND MASAYUKI YANO[1]

[1]University of Toronto Institute for Aerospace Studies
4925 Dufferin St; Toronto, ON, Canada; M3H 5T6
b.gibson@mail.utoronto.ca, masa.yano@utoronto.ca

Key words: Reduced Order Modelling, Hyperreduction, PDE-constrained Optimization, Trust Region Methods, Aerodynamic Shape Optimization

Abstract. We present a framework to accelerate optimization of problems where the objective function is governed by a nonlinear partial differential equation (PDE) using projection-based reduced-order models (ROMs) and a trust-region (TR) method. To reduce the cost of objective function evaluations by several orders of magnitude, we replace the underlying full-order model (FOM) with a series of hyperreduced ROMs (HROMs) constructed on-the-fly. Each HROM is equipped with an online-efficient *a posteriori* error estimator, which is used to define a TR. Hyperreduction is performed following a goal-oriented empirical quadrature procedure, which guarantees first-order consistency of the HROM with the FOM at the TR center. This ensures the optimizer converges to a local minimum of the underlying FOM problem. We demonstrate the framework through optimization of a nonlinear thermal fin and pressure-matching inverse design of an airfoil under Euler flow and Reynolds-averaged Navier-Stokes flow.

1 INTRODUCTION

Numerical approximations of partial differential equations (PDEs) play an increasingly prominent role in the development of new engineering designs. However, while higher-fidelity simulations of increasingly more complex systems become possible, the computational cost continues to grow. This motivates the current work, which attempts to preserve the accuracy of high-fidelity simulations while reducing the cost. Specifically, we wish to reduce the cost of optimization problems governed by general *nonlinear* PDEs. We employ a goal-oriented projection-based reduced-order model (ROM) constructed on-the-fly to achieve this goal.

Our approach to projection-based ROMs of *nonlinear* PDEs builds on three ingredients. The first ingredient is a reduced basis (RB) constructed from full-order model (FOM) solution snapshots. The RB provides a low-dimensional and rapidly converging approximation of the parameter-induced solution manifold [15]. The second ingredient is hyperreduction, which enables the rapid and accurate evaluation of the nonlinear residual. Of the many hyperreduction methods available [7, 5, 8, 3, 12, 12], we use the empirical quadrature procedure (EQP) [19, 23], which provides direct quantitative error control of the quantities of interest (i.e., goal-oriented error control) [22]. The combination of the RB and the EQP yields the hyperreduced ROM (HROM). The third ingredient is the dual-weighted residual (DWR) method [6], used in the context of ROM to approximate the error in the quantities of interest, as done in [16, 22].

ROM has been used extensively to accelerate PDE-constrained optimization problems [24, 1, 25, 20]. In these works, the ROM is built as the optimizer progresses using a trust-region (TR) approach, and is therefore referred to as "adaptive" or constructed "on-the-fly". There are also works that use a HROM constructed *a priori* (and not on-the-fly) in PDE-constrained optimization [2]. However, to the best of our knowledge, the present work is the first to employ on-the-fly hyperreduction and apply it to aerodynamic shape optimization.

The contributions of this work are threefold. Firstly, we develop a hyperreduction method for TR-based optimization which guarantees that the ROM satisfies the zeroth- and first-order consistency conditions required for global convergence. To achieve this goal, we develop the EQP with constraints tailored for gradient-based optimization. Secondly, we propose TR definitions informed by a DWR error estimate for nonlinear PDEs. Thirdly, we demonstrate the integrated framework to accelerate PDE-constrained optimization using a series of progressively more difficult optimization problems, including aerodynamic shape optimization.

2 FORMULATION

2.1 Problem statement

We introduce the optimization problem that we consider throughout this work. We first introduce an n_p-dimensional parameter domain $\mathcal{D} \subset \mathbb{R}^{n_p}$ and a parametrized d-dimensional physical domain $\Omega(\mu) \subset \mathbb{R}^d$. We then introduce a system of n_k conservation laws of the form

$$\nabla \cdot \big(F(u(\mu);\mu) + G(u(\mu), \nabla u(\mu);\mu)\big) = S(u(\mu), \nabla u(\mu);\mu) \quad \text{in } \Omega(\mu), \tag{1}$$

where $u(\mu) : \Omega(\mu) \to \mathbb{R}^{n_k}$ is the solution field, $F : \mathbb{R}^{n_k} \times \mathcal{D} \to \mathbb{R}^{d\times n_k}$ is the advection flux, $G : \mathbb{R}^{n_k} \times \mathbb{R}^{d\times n_k} \times \mathcal{D} \to \mathbb{R}^{d\times n_k}$ is the diffusion flux, and $S : \mathbb{R}^{n_k} \times \mathbb{R}^{d\times n_k} \times \mathcal{D} \to \mathbb{R}^{n_k}$ is the source. The system is complemented by appropriate boundary conditions. Given the state, we evaluate a functional output

$$q(u(\mu);\mu) \equiv \int_{\Omega(\mu)} q_V(u(\mu);\mu)\ dV + \int_{\partial\Omega(\mu)} q_B(u(\mu), \nabla u(\mu);\mu)\ dS, \tag{2}$$

where $q_V : \mathbb{R}^{n_k} \times \mathcal{D} \to \mathbb{R}$ and $q_B : \mathbb{R}^{n_k} \times \mathbb{R}^{d\times n_k} \times \mathcal{D} \to \mathbb{R}$ are the volume and boundary output integrands, respectively. We finally state our optimization problem: find μ_{opt} such that

$$\mu_{\text{opt}} = \arg\min_{\mu\in\mathcal{D}} C(q(u(\mu);\mu);\mu), \tag{3}$$

where $C(\cdot;\mu) : \mathbb{R} \to \mathbb{R}$ is a (low-cost, algebraic) function that maps the functional output to the objective function by, e.g., adding a penalty or comparing the output with a target value.

2.2 Finite element approximation

We now introduce a FOM using a finite element (FE) approximation of the optimization problem (3). To begin, we introduce a piecewise polynomial approximation space $\mathcal{V}_h$ of dimension $\mathcal{N}$, where $\mathcal{N}$ is typically large and is of order 10^4–10^6. We next introduce a semi-linear form $r_h : \mathcal{V}_h \times \mathcal{V}_h \times \mathcal{D} \to \mathbb{R}$ and a nonlinear form $q_h : \mathcal{V}_h \times \mathcal{D} \to \mathbb{R}$ associated with (1) and (2), respectively. The forms incorporate (i) stabilization terms required for non-conforming methods (e.g., discontinuous Galerkin (DG) method [4, 10]) and (ii) approximations of the integrals in

the weak form of (1) and (2) using piecewise Gauss(-like) quadrature rules with points $\{x_q\}_{q=1}^{N_q}$ and weights $\{\rho_q\}_{q=1}^{N_q}$. Note that N_q is the total number of volume and surface quadrature points, making it a mesh-dependent quantity; N_q is of order $\mathcal{N}$, but several times greater than $\mathcal{N}$. Our FE optimization problem is as follows: find $\mu_{\text{opt},h} \in \mathcal{D}$ such that

$$\mu_{\text{opt},h} = \arg\min_{\mu\in\mathcal{D}} C_h(\mu), \tag{4}$$

where $C_h(\mu) \equiv C(q_h(u_h(\mu);\mu);\mu)$ and $u_h(\mu) \in \mathcal{V}_h$ satisfies

$$r_h(u_h(\mu), v_h; \mu) = 0 \quad \forall v_h \in \mathcal{V}_h. \tag{5}$$

The solution of the optimization problem using a gradient-based optimization method (we employ an interior point method [17]) requires the evaluation of the gradient of $C_h(\cdot)$. We use an adjoint method to compute the gradient. To this end, we first introduce a dual problem: find $z_h(\mu) \in \mathcal{V}_h$ such that

$$r_h'[u_h(\mu)](v_h, z_h(\mu); \mu) = q_h'[u_h(\mu)](v_h;\mu) \quad \forall v_h \in \mathcal{V}_h, \tag{6}$$

where $r_h'[u_h(\mu)](w_h, v_h;\mu)$ and $q_h'[u_h(\mu)](w_h;\mu)$ are the Gâteaux derivatives of $r(\cdot, v_h;\mu)$ and $q_h(\cdot;\mu)$ about $u_h(\mu)$ in the direction w_h, respectively. We also define a residual sensitivity form $\partial_\mu r_h : \mathcal{V}_h \times \mathcal{V}_h \times \mathcal{D} \to \mathbb{R}^{n_p}$ and an output sensitivity form $\partial_\mu q_h : \mathcal{V}_h \times \mathcal{D} \to \mathbb{R}^{n_p}$ such that $\partial_\mu r_h(w_h, v_h;\mu) \equiv \frac{\partial}{\partial\mu} r_h(w_h, v_h;\mu)$ and $\partial_\mu q_h(v_h;\mu) \equiv \frac{\partial}{\partial\mu} q_h(v_h;\mu)$ for all $w_h, v_h \in \mathcal{V}_h$ and $\mu \in \mathcal{D}$. We then express the gradient of $q_h(u_h(\cdot);\cdot)$ as

$$\nabla_\mu(q_h(u_h(\mu);\mu)) = -\partial_\mu r_h(u_h(\mu), z_h(\mu);\mu) + \partial_\mu q_h(u_h(\mu);\mu). \tag{7}$$

The implicit dependence on μ through u_h is captured by the first term, and the direct dependence through q_h is captured by the second term. Given $\nabla_\mu(q_h(u_h(\mu);\mu))$, we appeal to the chain rule to evaluate the objective function gradient $\nabla_\mu C_h(\mu) \in \mathbb{R}^{n_p}$.

The evaluation of the objective $C_h(\mu)$ and the associated gradient $\nabla_\mu C_h(\mu)$ requires the solution of the FE primal problem (5) and the dual problem (6). This requires $\mathcal{O}(\mathcal{N})$ operations, which renders the solution of the optimization problem (3) computationally expensive.

2.3 Reduced-order model

We now introduce a projection-based ROM, which we use as a surrogate model to rapidly approximate the objective function $C_h(\mu)$ and its gradient $\nabla_\mu C_h(\mu)$, accelerating the solution of the optimization problem using a TR method. To ensure the global convergence of the TR method, the ROM approximations $C_N(\mu)$ and $\nabla_\mu C_N(\mu)$ must satisfy the following two conditions:

- Zero-order consistency (ZOC): $C_N(\mu_{\text{TR}}) = C_h(\mu_{\text{TR}})$,
- First-order consistency (FOC): $\nabla_\mu C_N(\mu_{\text{TR}}) = \nabla_\mu C_h(\mu_{\text{TR}})$,

where $\mu_{\text{TR}} \in \mathcal{D}$ is the current TR center. If these two conditions are satisfied, and each subproblem satisfies a sufficient decrease condition, then the TR method is guaranteed to converge to a local minimum of the optimization problem (4) [17, 24]. We construct the ROM with these conditions in mind.

2.3.1 Reduced basis method

We now describe on-the-fly construction of the RB *in the context of TR-based optimization*. The RB method takes advantage of the fact that the parametric solution manifold $\{u_h(\mu)\}_{\mu\in\mathcal{D}}$ is often amenable to approximation in a lower-dimensional space than the FE space $\mathcal{V}_h$. In the context of optimization, we obtain the RB using Gram-Schmidt orthogonalization (GSO) (alternatively POD) on the FOM solutions obtained so far during optimization. Namely, we choose the set of all parameter values at which the FOM problems (5) and (6) have already been solved as the training set $\Xi_{\text{train}} \subset \mathcal{D}$ (which includes μ_{TR}) and obtain the associated (primal) training snapshots $U_{\text{train}} \equiv \{u_h(\mu_t)\}_{\mu_t\in\Xi_{\text{train}}}$. We then set $\{\zeta_i^{\text{pr}}\}_{i=1}^{N_{\text{pr}}} = \text{GSO}(U_{\text{train}} \cup z_h(\mu_{\text{TR}}))$ and finally introduce the RB space $\mathcal{V}_N^{\text{pr}} \equiv \text{span}\{\zeta_i^{\text{pr}}\}_{i=1}^{N_{\text{pr}}}$ of dimension $N_{\text{pr}} \ll \mathcal{N}$. See Remark 1 for the rationale in the choice of the particular RB space.

Given the RB space, our RB approximation of the primal problem (5) is the following: given $\mu \in \mathcal{D}$, find $u_N(\mu) \in \mathcal{V}_N^{\text{pr}}$ such that

$$r_h(u_N(\mu), v; \mu) = 0 \quad \forall v \in \mathcal{V}_N^{\text{pr}}. \tag{8}$$

We then evaluate the output $C_N(\mu) \equiv C(q_h(u_N(\mu);\mu);\mu)$. To evaluate the gradient $\nabla_\mu C_N(\mu)$ using the the adjoint method, we first introduce the dual problem: find $z_N^{\text{pr}}(\mu) \in \mathcal{V}_N^{\text{pr}}$ such that

$$r_h'[u_N(\mu)](v_N, z_N^{\text{pr}}(\mu);\mu) = q_h'[u_N(\mu)](v_N;\mu) \quad \forall v_N \in \mathcal{V}_N^{\text{pr}}. \tag{9}$$

Then we exactly evaluate $\nabla_\mu(q_h(u_h(\mu);\mu))$ and $\nabla_\mu C_N(\mu)$ using a procedure analogous to that described in Section 2.2, with $u_h(\mu)$ and $z_h(\mu)$ replaced by $u_N(\mu)$ and $z_N^{\text{pr}}(\mu)$.

Remark 1. The inclusion of a solution $u_h(\mu_{\text{TR}})$ in the training set U_{train} ensures $u_N(\mu_{\text{TR}}) = u_h(\mu_{\text{TR}})$: i.e., we exactly reproduce primal solutions. Hence, we satisfy the ZOC condition: $C_N(\mu_{\text{TR}}) = C_h(\mu_{\text{TR}})$. Similarly, the dual solution $z_h(\mu_{\text{TR}})$ is included in the primal RB set to ensure that (9) yields the exact solution: $z_N^{\text{pr}}(\mu_{\text{TR}}) = z_h(\mu_{\text{TR}})$. Together with the fact that $u_N(\mu_{\text{TR}}) = u_h(\mu_{\text{TR}})$, this ensures that the ROM satisfies the FOC condition: $\nabla_\mu C_N(\mu_{\text{TR}}) = \nabla_\mu C_h(\mu_{\text{TR}})$.

2.3.2 Hyperreduction by an empirical quadrature procedure

Despite the significant reduction in the dimension of the approximation space (i.e., $N_{\text{pr}} \ll \mathcal{N}$), the cost to solve (8) for $u_N(\mu) \in \mathcal{V}_N^{\text{pr}}$ as described above still scales with $\mathcal{N}$ because the evaluation of the residual scales with the number of FE quadrature points $N_q = \mathcal{O}(\mathcal{N})$. The goal of hyperreduction is to achieve $\mathcal{N}$-independent cost.

While many hyperreduction methods exist, we use the EQP [23]. The EQP seeks a reduced quadrature (RQ) rule $\{(\tilde{x}_q, \tilde{\rho}_q)\}_{q=1}^{\tilde{N}_q}$, where $\tilde{N}_q \ll N_q$ to promote rapid evaluation while controlling the quadrature error in integrals that are relevant to the ROM. Accuracy is promoted by training on a set of integrals $\{I_t \equiv \int_\Omega \mathcal{I}_t(x)dx\}_{t=1}^{n_{\text{EQP}}}$ associated with integrands $\mathcal{I}_t : \Omega \to \mathbb{R}$, where each I_t is evaluated using the FE quadrature rule $\{(x_q, \rho_q)\}_{q=1}^{N_q}$. We then pose the following optimization problem: find $\hat{\rho} \in \mathbb{R}_{\geq 0}^{N_q}$ such that

$$\hat{\rho} = \underset{\rho \in \mathbb{R}^{N_q}_{\geq 0}}{\arg\min} \|\rho\|_0 \tag{10}$$
$$\text{subject to} \quad \left| \sum_{q=1}^{N_q} \rho_q \mathcal{I}_t(x_q) - I_t \right| \leq \epsilon_{\text{EQP},t}, \quad t = 1, \ldots, n_{\text{EQP}},$$

where $\epsilon_{\text{EQP},t} \in \mathbb{R}_{\geq 0}$ is a user-prescribe tolerance for the t-th training integral. We then extract non-zero entries of $\hat{\rho}$ to obtain the RQ rule $\{(\tilde{x}_{\tilde{q}}, \tilde{\rho}_{\tilde{q}})\}_{\tilde{q}=1}^{\tilde{N}_q} = \{(x_q, \hat{\rho}_q) \mid \hat{\rho}_q \neq 0, q = 1, ..., N_q\}$.

Given the RQ rule, we evaluate the objective $\tilde{C}_N(\mu)$ and its gradient $\nabla \tilde{C}_N(\mu)$ associated with the hyperreduced ROM (HROM) as follows. We find $\tilde{u}_N(\mu) \in \mathcal{V}_N^{\text{pr}}$ such that

$$\tilde{r}_h(\tilde{u}_N(\mu), v_N; \mu) = 0 \quad \forall v_N \in \mathcal{V}_N^{\text{pr}}, \tag{11}$$

where $\tilde{r}_h(\cdot, \cdot; \mu)$ is the approximation of the semi-linear form $r_h(\cdot, \cdot; \mu)$ using the RQ rule. We then evaluate the HROM approximation of the objective, $\tilde{C}_N(\mu) \equiv C(q_h(\tilde{u}_N(\mu); \mu); \mu)$. In this work, we do not hyperreduce the output functional $q_h(\cdot; \mu)$, since the cost of its evaluation is relatively low. To evaluate its gradient, we first find $\tilde{z}_N^{\text{pr}}(\mu) \in \mathcal{V}^{\text{pr}}$ such that

$$\tilde{r}_h'[\tilde{u}_N(\mu)](v_n, \tilde{z}_N^{\text{pr}}(\mu); \mu) = q_h'[\tilde{u}_N(\mu)](v_N; \mu) \quad \forall v_N \in \mathcal{V}_N^{\text{pr}}, \tag{12}$$

and then follow the procedure in Section 2.2, with $u_N(\mu)$ and $z_N^{\text{pr}}(\mu)$ replaced by $\tilde{u}_N(\mu)$ and $\tilde{z}_N^{\text{pr}}(\mu)$, to obtain $\nabla_\mu \tilde{C}_N(\mu)$.

We now describe the key to our hyperreduction procedure: the selection of the training integrals in the EQP. In the EQP, we are free to choose the training integrals, which enables construction of goal-oriented, application-tailored empirical quadrature rules. Our application is optimization, so we use constraints that preserve ZOC and FOC of the surrogate model:

$$\left| r_h(u_N(\mu_{\text{TR}}), \zeta_i^{\text{pr}}; \mu_{\text{TR}}) - \tilde{r}_h(u_N(\mu_{\text{TR}}), \zeta_i^{\text{pr}}; \mu_{\text{TR}}) \right| \leq \epsilon_{\text{FOC}}, \ i = 1, ..., N_{\text{pr}}, \tag{13}$$
$$\left| r_h'[u_N(\mu_{\text{TR}})](\zeta_i^{\text{pr}}, z_N^{\text{pr}}(\mu_{\text{TR}}); \mu_{\text{TR}}) - \tilde{r}_h'[u_N(\mu_{\text{TR}})](\zeta_i^{\text{pr}}, z_N^{\text{pr}}(\mu_{\text{TR}}); \mu_{\text{TR}}) \right| \leq \epsilon_{\text{FOC}}, \ i = 1, ..., N_{\text{pr}}, \tag{14}$$
$$\left| \partial_\mu r_h(u_N(\mu_{\text{TR}}), z_N^{\text{pr}}(\mu_{\text{TR}}); \mu_{\text{TR}}) - \partial_\mu \tilde{r}_h(u_N(\mu_{\text{TR}}), z_N^{\text{pr}}(\mu_{\text{TR}}); \mu_{\text{TR}}) \right| \leq \epsilon_{\text{FOC}}, \tag{15}$$

where we set ϵ_{FOC} to be near machine precision, so that the integrals are (nearly) exactly reproduced by the RQ rule.

Remark 2. Constraint (13) ensures that $\tilde{u}_N(\mu_{\text{TR}}) = u_N(\mu_{\text{TR}})$, constraint (14) ensures that $\tilde{z}_N^{\text{pr}}(\mu_{\text{TR}}) = z_N^{\text{pr}}(\mu_{\text{TR}})$, and constraint (15) ensures that $\nabla_\mu(q(\tilde{u}_N(\mu); \mu))|_{\mu_{\text{TR}}} = \nabla_\mu(q(u_N(\mu); \mu))|_{\mu_{\text{TR}}}$. This ensures that ZOC and FOC is carried from the FOM through ROM to the final HROM: i.e., $C_h(\mu_{\text{TR}}) = C_N(\mu_{\text{TR}}) = \tilde{C}_N(\mu_{\text{TR}})$ and $\nabla_\mu C_h(\mu_{\text{TR}}) = \nabla_\mu C_N(\mu_{\text{TR}}) = \nabla_\mu \tilde{C}_N(\mu_{\text{TR}})$. The total number of constraints in the EQP is $2N_{\text{pr}} + n_p$. We may also add more constraints at a lighter tolerance to improve the quality of the empirical quadrature rule.

2.3.3 *A posteriori* error estimation

Following [24, 25, 20], we wish to use TRs informed by the ROM error. Since we consider general nonlinear PDEs in this work, we use the DWR method [6], following the construction of

a DWR-based error estimate for HROMs in [22]. Namely, given the dual FE snapshots $Z_{\text{train}} \equiv \{z_h(\mu_t)\}_{\mu_t \in \Xi_{\text{train}}}$, we first construct a dual RB $\{\zeta_i^{\text{du}}\}_{i=1}^{N_{\text{du}}} = \text{GSO}(Z_{\text{train}})$ and the associated dual RB space $\mathcal{V}_N^{\text{du}} \equiv \text{span}\{\zeta_i^{\text{du}}\}_{i=1}^{N_{\text{du}}}$. We then find the dual RB solution $z_N^{\text{du}}(\mu) \in \mathcal{V}_N^{\text{du}}$ such that

$$r_h'[\tilde{u}_N(\mu)](v_N, z_N^{\text{du}}(\mu); \mu) = q_h'[\tilde{u}_N(\mu)](v_N; \mu) \quad \forall v_N \in \mathcal{V}_N^{\text{du}} \tag{16}$$

and estimate the error incurred by the HROM (with respect to FOM) as

$$|q(u_h(\mu); \mu) - q(\tilde{u}_N(\mu); \mu)| \approx \eta_N(\mu) \equiv |r_h(\tilde{u}_N(\mu), z_N^{\text{du}}(\mu); \mu)|. \tag{17}$$

In this work (unlike in [22]), we do not apply hyperreduction to the dual problem; in the context of optimization, we find that the cost to construct the RQ rule for the dual problem using the EQP outweighs the acceleration the dual HROM provides.

2.4 TR definitions

In general, the kth trust region $\mathcal{D}_{\text{TR},k} \subset \mathcal{D}$ is defined as $\mathcal{D}_{\text{TR},k} \equiv \{\mu \in \mathcal{D} \mid \theta_k(\mu) \leq \delta_k\}$, where $\theta_k : \mathcal{D} \to \mathbb{R}_{\geq 0}$ is the TR metric, and δ_k is the maximum allowable TR metric (or TR size). In this work, we terminate solution of the optimization sub-problem when the optimizer reaches $\theta_k(\mu) \geq 0.9\delta_k$, or within 90% of the TR boundary. We consider three TR metrics.

Geometric TR (GTR). In the standard GTR, $\theta_k^{\text{GTR}}(\mu) \equiv \|\mu - \mu_{\text{TR}}\|_2$, the Euclidean distance between the current point μ and the TR center μ_{TR}. The TR size δ_k adapts based on the accuracy of the previous HROM relative to the FOM after a model update [17].

Error-based TR (ETR). The GTR does not directly capture the accuracy of the HROM, especially when the design parameters are scaled poorly. The HROM error is more accurately predicted by the error estimate $\eta_N(\mu)$ in (17). We therefore define an ETR with a metric $\theta_k^{\text{ETR}}(\mu) \equiv \eta_{N=k}(\mu)$. To ensure the TR is bounded, we also use a large maximum GTR with a fixed size $\mathcal{D}_{\text{maxGTR}}$ to bound the ETR, so $\mathcal{D}_{\text{ETR}} \equiv \{\mu \in \mathcal{D} \mid \theta_k^{\text{ETR}}(\mu) \leq \delta_k\} \cap \mathcal{D}_{\text{maxGTR}}$.

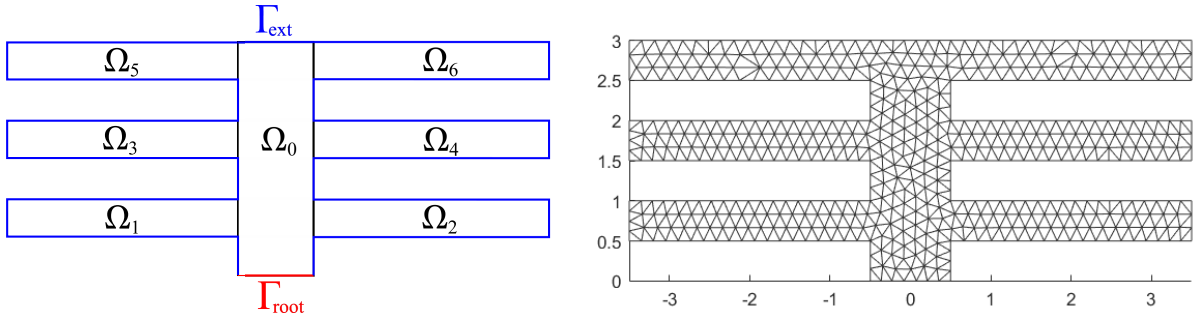
Advanced TR (ATR). The ATR uses the error estimate to attempt to guarantee that the newly selected design point satisfies the sufficient decrease condition. If the error is predicted to be large, but the decrease in the objective function is suspected to be even larger, then we accept the step. This is because typically near the beginning, the HROM is not that accurate, but it also does not need to be as accurate in order to make progress. Hence we choose $\theta_k^{\text{ATR}}(\mu) \equiv \eta_{N=k}(\mu)\big((\tilde{C}_k(\mu_{\text{TR}}) - \tilde{C}_k(\mu)) + \beta(\mu - \mu_{\text{TR}})^T \nabla \tilde{C}_k(\mu_{\text{TR}})\big)^{-1}$ and fix the TR size to $\delta = 1$. To ensure the TR is neither unbounded nor empty, we also use a small minimum ETR and a maximum GTR: $\mathcal{D}_{\text{ATR}} \equiv \big(\{\mu \in \mathcal{D} \mid \theta_{\text{ATR}}(\mu) \leq 1\} \cup \mathcal{D}_{\text{minETR}}\big) \cap \mathcal{D}_{\text{maxGTR}}$.

3 RESULTS

We assess the proposed ROM-accelerated optimization method using three model problems. Due to space limitations, we only present a sample of the results here; for more details, see [13].

3.1 Nonlinear thermal fin

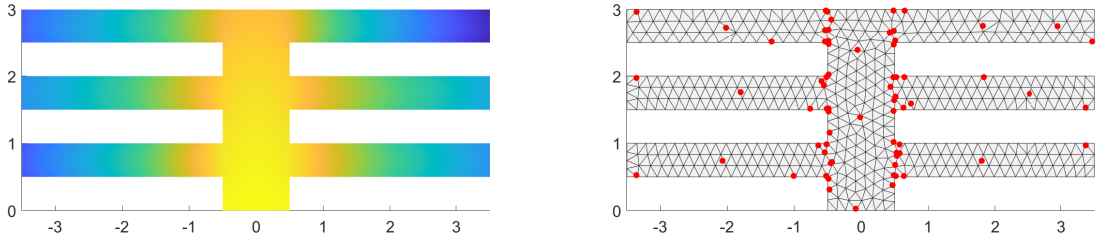
We consider a $n_p = 7$-dimensional nonlinear thermal fin problem, modelled on Qian et al. [20], but governed by a *nonlinear* heat equation to motivate hyperreduction. Figure 1 shows the physical domain Ω. The conductivity of the main trunk Ω_0 is $e^{u(\mu)}$, and the conductivity of each



(a) Designation of regions and boundaries, Γ_{root} in red and Γ_{ext} in blue

(b) Mesh for FOM ($\mathcal{N} = 2{,}029$ degrees of freedom)

Figure 1: Model thermal fin problem



(a) ζ_1^{pr}

(b) Mesh overlayed with quadrature points (red)

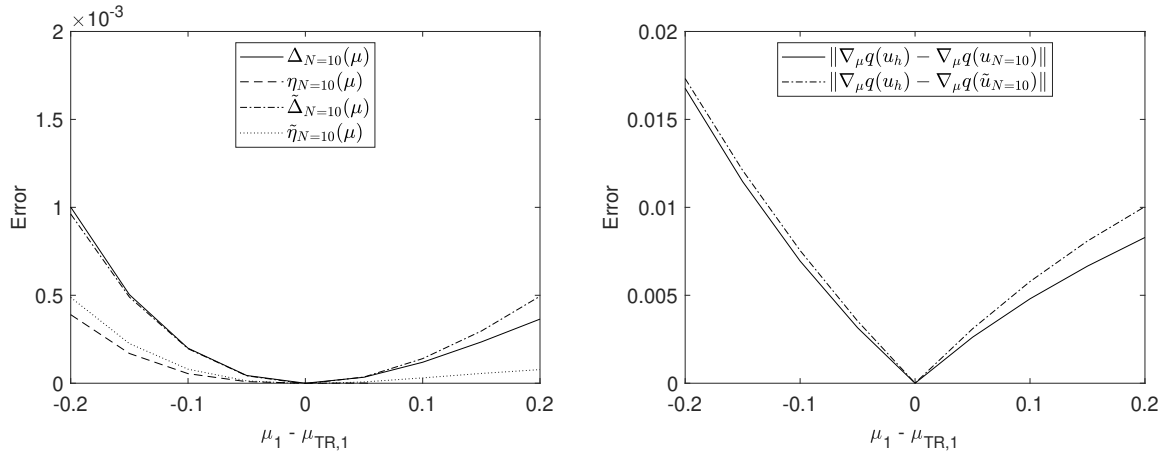
Figure 2: RB and RQ visualization

fin Ω_i is $10^{\mu_i} e^{u(\mu)}$, $i = 1, ..., 6$. The temperature-dependent conductivity makes the heat equation nonlinear. Natural convection away from Γ_{ext} is described by $\text{Bi}(\mu) = 10^{\mu_7}$. The objective is to minimize the (scaled) average temperature $\int_\Omega u(\mu) dx$, with a penalty on higher values of each μ_i.

Figure 2a shows the first RB function. The RB is computed using POD (but without compression), making ζ_1^{pr} the average solution of $(U_{\text{train}} \cup z_h(\mu_{\text{TR}}))$. Figure 2b shows a RQ rule, where $N_q = 5{,}256$ is reduced to $\tilde{N}_q = 69$. Most RQ points are clustered near the subdomain boundaries.

Figure 3 demonstrates the ZOC and FOC of the model for $N = 10$. Both the error $\Delta(\mu) \equiv |q(u_h(\mu); \mu) - q(u_N(\mu); \mu)|$ and the error estimate $\eta_N(\mu)$ are shown to be zero at μ_{TR}, for both the non-hyperreduced ROM and the HROM. The error in the gradient of the objective is also zero, demonstrating the successful guarantee of ZOC and FOC; cf. Remarks 1 and 2.

Table 1 breaks down the cost to solve the optimization problem for the three TR definitions. Speedup is observed for all three cases, though most significantly with the GTR. This is due to fewer FOM solves required and not having to evaluate $\eta_N(\mu)$. The parameterization of this problem is not that complex, so the advantages of the error-based TRs over a GTR with an adaptive TR size update rule are not clear.



(a) Output error (b) Gradient error

Figure 3: Comparison of non-hyperreduced and HROM output and gradient errors near μ_{TR}, with $N = 10$

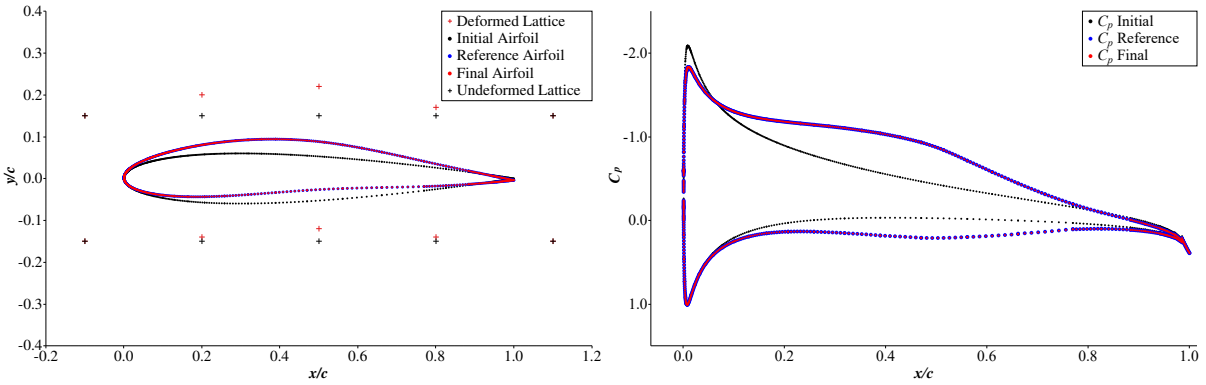
Table 1: Results for nonlinear thermal fin optimization

Case	**# FOM**	**# HROM**	$t_{\textbf{FOM}}$ [s]	$t_{\textbf{HROM}}$ [s]	t_η [s]	$t_{\textbf{train}}$ [s]	$t_{\textbf{total}}$ [s]
FOM	16	-	6.4	-	-	-	6.4
GTR	11	60	4.0	0.1	-	0.2	4.6
ETR	12	79	4.5	0.2	0.5	0.3	6.0
ATR	12	59	4.6	0.1	0.4	0.3	6.0

3.2 Inverse design: Euler equations

We consider a pressure-matching inverse-design problem governed by the Euler equations. To this end, we consider free-form deformation (FFD) of the NACA0012 airfoil. The reference airfoil surface pressure p_{ref} is associated with a reference deformation μ_{ref}. Our goal is to reproduce μ_{ref} by optimizing the objective function $q(u(\mu);\mu) \equiv \int_{\Gamma_{\text{foil}}} (p(u(\mu)) - p_{\text{ref}})^2 dS$. This is an inverse design problem — finding the geometry that yields a specified pressure distribution. Zahr and Farhat [25] studied this problem using ROM, but without hyperreduction and *a posteriori* error estimates for the output. Figure 4 shows the geometry and surface pressure distributions; the reference and final geometries are indistinguishable. Figure 4a also shows the FFD control lattice, characterized by $n_p = 6$ parameters. The FOM uses a high-order DG FEM with output-based adaptive mesh refinement [6], with $\mathcal{N} = 20{,}496$. The use of these state-of-the-art techniques ensures that the further cost reduction obtained by the ROM-accelerated methods provides a fair comparison, not skewed by the choice of an inefficient baseline.

We solve the problem using the HROM-accelerated method and the ATR. We also experiment with a parameter in the algorithm called n_{warmup}, which is the number of initial FOM solves (after the first one, which is not optional) which are performed before starting to apply HROM-acceleration. The motivation for this is that in the beginning with as few as only one



(a) Geometry and FFD parametrization

(b) Reference C_p distribution

Figure 4: Geometry and C_p distribution for Euler-based pressure matching optimization, with $M = 0.2$ and $\alpha = 5°$

Table 2: Results for Euler inverse aerodynamic design

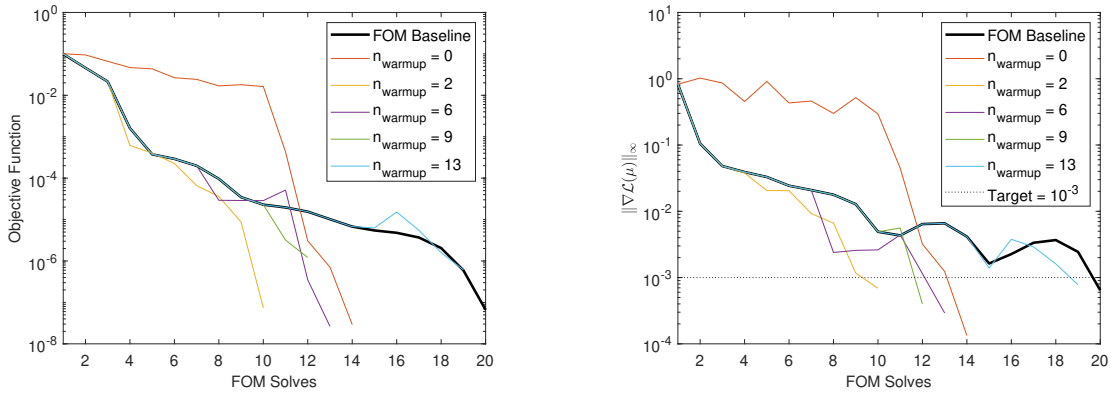
Case	**# FOM**	**# HROM**	$t_{\mathbf{FOM}}$ [s]	$t_{\mathbf{HROM}}$ [s]	t_η [s]	$t_{\mathbf{train}}$ [s]	$t_{\mathbf{total}}$ [s]
FOM	20	-	33.6	-	-	-	33.6
$n_{\text{warmup}} = 0$	14	198	25.3	1.9	0.9	10.1	38.1
$n_{\text{warmup}} = 2$	10	62	17.8	0.5	0.3	2.6	21.3
$n_{\text{warmup}} = 6$	13	46	23.6	0.4	0.3	7.3	31.6
$n_{\text{warmup}} = 9$	12	14	21.9	0.1	0.1	3.1	25.2
$n_{\text{warmup}} = 13$	19	73	33.2	0.9	0.5	64.4	99.1

training snapshot, the resulting ROM is quite crude, perhaps resulting in more exploration of the parameter space and therefore more FOM solves. The convergence in Figure 5 and the timing breakdown in Table 2 demonstrate the effect of n_{warmup} on performance, as compared to the FOM baseline. All cases except the two extremes of $n_{\text{warmup}} = 0$ and 13 show speedup. The initial HROM is too crude for $n_{\text{warmup}} = 0$ which impedes the initial progress, and more warmup solves than necessary are performed for $n_{\text{warmup}} = 13$.

3.3 Inverse Design: RANS Equations

We solve another inverse design problem similar to the one in Section 3.2 with two modifications: the governing equations are the Reynolds-averaged Navier-Stokes (RANS) equations, and the angle of attack is a design variable. The FOM has $\mathcal{N} = 77{,}940$ DOF. The pressure distribution is similar to Figure 4b, so we do not show it here. As far as we know, a RANS-based optimization problem accelerated by ROM has not been attempted in the literature; we wish to push the boundaries of what has been achieved to date with ROM-accelerated optimization.

Unfortunately, as shown in Table 3, actual acceleration in terms of computational time is not observed. We reduce the number of FOM solves, but we do not see speedup in real terms. This is because of the high cost both of the HROM solves and of the training, specifically the EQP



(a) Objective function

(b) First order optimality

Figure 5: Effect of number of initial FOM solves n_{warmup} on convergence of the Euler pressure matching problem using HROM acceleration with an ATR

Table 3: Results for RANS inverse aerodynamic design, $\text{Re} = 2 \times 10^5$, $\epsilon_{\text{opt}} = 10^{-3}$

Case	**# FOM**	**# HROM**	$t_{\textbf{FOM}}$ [s]	$t_{\textbf{HROM}}$ [s]	t_η [s]	$t_{\textbf{train}}$ [s]	$t_{\textbf{total}}$ [s]
FOM	21	-	341.3	-	-	-	341.3
GTR	16	91	272.3	49.5	-	122.6	444.9
ETR	13	127	222.5	69.4	12.3	64.5	369.1
ATR	14	142	246.2	81.0	14.1	127.6	469.1

training. HROM solves are expensive not because of any failure in the method described here, but simply due to the expense of the FFD transformation, which we were not able to reduce. We still require a full geometry transformation, the cost of which is still mesh-dependent. The EQP training cost also scales with N_q. More work is required to reduce both of these contributions to the cost before we can see true speedup.

4 CONCLUSIONS

We have presented a framework to accelerate nonlinear PDE-constrained optimization using a projection-based HROM and TR method. The HROM is constructed to achieve ZOC and FOC with the FOM at the TR center, providing an optimization convergence guarantee. We use the DWR error estimator $\eta_N(\mu)$ to describe two error-aware TRs, in addition to the standard GTR. We then present results on three model problems, two of which did see a reduction in computational cost. The third problem represents the breaking of new ground in ROM-accelerated optimization, but to date real computational speedup has not yet been achieved.

Future work on this subject should involve further reducing the cost of the FFD transformation and the EQP. Additionally, the stability of the HROM for RANS problems needs to be improved; we have encountered cases where the nonlinear HROM solver failed to converge. Finally, more and more complex optimization problems should be attempted, until ROM-accelerated methods are able to handle industry-scale aerodynamic shape optimization problems.

ACKNOWLEDGMENTS

This work was financially supported by the Natural Science and Engineering Research Council of Canada through an Alexander Graham Bell Canada Graduate Scholarship and Discovery Grant. Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation, the Government of Ontario, Ontario Research Fund - Research Excellence, and the University of Toronto.

REFERENCES

[1] Agarwal, A. and Biegler, L. A trust-region framework for constrained optimization using reduced order modeling. *Optimization in Engineering* (2011) **37**:3–35.

[2] Amsallem, D., Zahr, M., Choi, Y., and Farhat, C. Design optimization using hyper-reduced-order models. *Struct. Multidiscip. Optim.* (2015) **51**:919–940.

[3] An, S. S., Kim, T., and James, D. L. Optimizing cubature for efficient integration of subspace deformations. *ACM Trans. Graph.* (2008) **27**:165:1–165:10.

[4] Arnold, D. N., Brezzi, F., Cockburn, B., and Marini L.D. Unified analysis of discontinuous Galerkin methods for elliptical problems. *SIAM Journal on Numerical Analysis* (2002) **39**:1749–1779.

[5] Barrault, M., Maday, Y., Nguyen, N.C., and Patera, A.T. An "empirical interpolation" method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris, Ser. I* (2004) **339**:667–672.

[6] Becker, R. and Rannacher, R. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica* (2001) **10**:1–102. Cambridge University Press.

[7] Bui-Thanh, T., Damodaran, M., and Willcox, K. Proper orthogonal decomposition extensions for parametric applications in compressible aerodynamics. *21st AIAA Applied Aerodynamics Conference* (2003)

[8] Carlberg, K., Bou-Mosleh, C., and Farhat, C., Efficient non-linear model reduction via a least-squares Petrov-Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering* (2011) **86**:155-181.

[9] Chaturantabut, C. and Sorensen, D.C. Nonlinear model reduction via Discrete Empirical Interpolation. *SIAM Journal on Scientific Computing* (2010) **32**:2737–2764.

[10] Cockburn, B. Discontinuous Galerkin methods. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* (2003) 83:731–754.

[11] Everson, R. and Sirovich, L. Karhunen-Loève procedure for gappy data. *Journal of the Optical Society of America A, Optics and Image Science* (1995) **12**:1657–1664.

[12] Farhat, C., Avery, P., Chapman, T., and Cortial, J. Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and

weighting for computational efficiency. *International Journal for Numerical Methods in Engineering* (2014) **98**:625–662.

[13] Gibson, B. Accelerated PDE-Constrained Optimization by Adaptive Reduced Order Modelling and Goal-Oriented Hyperreduction. *University of Toronto* (2022). Master's Thesis.

[14] Gogu, C. Improving the efficiency of large scale topology optimization through on-the-fly reduced order model construction. *International Journal for Numerical Methods in Engineering* (2015) **101**:281–304

[15] Hesthaven, J.S., Rozza, G., and Stamm, B. *Certified reduced basis methods for parametrized partial differential equations*. Springer (2016).

[16] Meyer, M. and Matthies, H.G. Efficient model reduction in non-linear dynamics using the Karhunen-Loève expansion and dual-weighted-residual methods. *Computational Mechanics* (2003) **31**:179–191

[17] Nocedal, J. and Wright, S. *Numerical optimization*. Springer-Verlag GmbH (2006).

[18] Patera, A.T. and Rozza, G. *Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations*. MIT Pappalardo Graduate Monographs in Mechanical Engineering (2006).

[19] Patera, A.T. and Yano, M. An LP empirical quadrature procedure for parametrized functions. *Comptes Rendus Mathematique* (2017) **355**:1161–1167.

[20] Qian, E., Grepl, M., Veroy, K., and Willcox, K. A certified trust region reduced basis approach to PDE-constrained optimization. *SIAM Journal on Scientific Computing* (2017) **39**:S434–S460.

[21] Yano, M. Discontinuous Galerkin reduced basis empirical quadrature procedure for model reduction of parametrized nonlinear conservation laws. *Advances in Computational Mathematics* (2019) **45**:2287–2320.

[22] Yano, M. Goal-oriented model reduction of parametrized nonlinear PDEs; application to aerodynamics. *International Journal for Numerical Methods in Engineering* (2020) (accepted).

[23] Yano, M. and Patera, A.T. An LP empirical quadrature procedure for reduced basis treatment of parametrized nonlinear PDEs. *Computer Methods in Applied Mechanics and Engineering* (2019) **344**:1104–1123.

[24] Yue, Y. and Meerbergen, K. Accelerating optimization of parametric linear systems by model order reduction. *Society for Industrial and Applied Mathematics Journal of Optimization* (2013) **23**:1344–1370.

[25] Zahr, M.J. and Farhat, C. Progressive construction of a parametric reduced-order model for PDE-constrained optimization. *International Journal for Numerical Methods in Engineering* (2015) **102**:1111–1135.