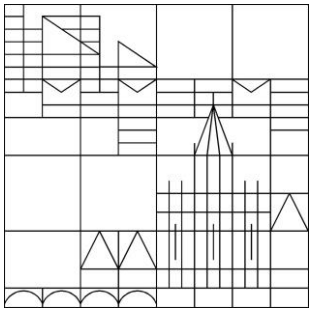


Universität Konstanz

Reduced basis methods for optimal experimental design of parametrized linear evolution problems

Andrea Petrocchi
Matthias K. Scharrer
Stefan Volkwein

Konstanzer Schriften in Mathematik

Nr. 403, März 2022

ISSN 1430-3558

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1l9b7uuht7b1s0

© Fachbereich Mathematik und Statistik
Universität Konstanz
Fach D 197, 78457 Konstanz, Germany

Reduced Basis Methods for Optimal Experimental Design of Parametrized Linear Evolution Problems

Andrea Petrocchi * **Matthias K. Scharrer** ** **Stefan Volkwein** *

* *Department of Mathematics and Statistics, Universität Konstanz, Universitätsstr. 10, D-78457 Konstanz, Germany. (e-mail: andrea.petrocchi@uni-konstanz.de, stefan.volkwein@uni-konstanz.de)*
** *Virtual Vehicle Research GmbH, Inffeldgasse 21A, A-8010 Graz, Austria. (e-mail: matthias.scharrer@v2c2.at)*

Abstract: In this paper we propose an algorithm for the bi-level optimal experimental design involving a parameter-dependent evolution problems. In the inner cycle a control is fixed and the parameter is optimized in order to minimize a cost function that measure the discrepancy from some data. In the outer cycle the found parameter is fixed and the control is now optimized in order to minimize a suitable measure of uncertainty of the parameters. The inner cycle uses a trust-region reduced basis approximation of the model with creation and enrichment of the reduced basis on-the-fly. Numerical examples illustrate the efficiency of the proposed approach.

Keywords: Reduced basis methods, a-posteriori error, trust-region optimization, evolution problems, optimal experimental design.

1. INTRODUCTION

PDE modelling is relevant in many fields, such as medicine and engineering, and optimization methods are useful when dealing with real-life problems. It is not uncommon for some models to depend on underlying unknown parameters, while we might have some controls (either finite or infinite-dimensional) that we are able to change in real time. Parameter estimation (PE) comprehends methods and algorithms able to find or approximate the underlying parameters using empirical data or observations. It is often the case that the choice of a particular control speeds up or slows down the parameter estimation, hence being an important choice when the temporal or computational costs are limited.

In this work two problems are discussed. First, the step of parameter estimation is not trivial: usually such problems are ill-posed or non-convex, hence resulting in non-unique or non-global solutions. Even when well-posed, some applications deal with such high-dimensional discretizations that the solution is prohibitively costly. For this reason model order reduction (MOR) methods were developed, where the "expensive" models are replaced by cheaper and less accurate surrogates; cf., e.g., Benner et al. (2020a,b,c) and Benner et al. (2014). In our case the reduced basis (RB) is used, creating a reduced-order model on the optimization path in such a way that the local minimum found by the optimization method in the reduced space is very close to a local minimum in the full space.

The second problem belongs to the optimal experimental design (OED); see e.g. Alexanderian et al. (2016); Bock et al. (2013); Körkel et al. (2004); Pozzi et al. (2019); Reddy et al. (2019). We want to find the "best" control, namely the one that gives us a better parameter estimation. In particular, assuming there are measurement error in the observations, the "best" control is the one that minimizes the uncertainty given by the parameter optimization.

The article is organized as follows: In Section 2 the parameter estimation problem is introduced. Its discretization and the a-posteriori error analysis is discussed in Section 3. The trust-region method for the parameter optimization is explained in Section 4. The OED is described in Section 5, whereas Section 6 is devoted for numerical experiments.

2. PARAMETER ESTIMATION PROBLEM

We assume that all parameters are stacked in a vector $\mu \in \mathbb{R}^d$, and belong to the compact (admissible) set

$$\mathcal{P}_{\mathsf{ad}} = \left\{ \mu \in \mathbb{R}^d \,\middle|\, \mu_i^{\mathsf{a}} \le \mu_i \le \mu_i^{\mathsf{b}} \text{ for } i = 1, \dots, d \right\}$$

The *state* variable y satisfies the evolution problem

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} \langle y(t), \varphi \rangle_H + a_\mu(y(t), \varphi) &= \langle f_\mu(t; u), \varphi \rangle_{V',V} \\ y(0) &= y_\circ \quad \text{in } H \end{aligned} \tag{1}$$

for all $\varphi \in V$ and $t \in (0, T]$ a.a., where $y_\circ \in H$, V, H are Hilbert spaces with $V \hookrightarrow H \hookrightarrow V'$ (Gelfand triple) and u denotes the *control* belonging to a convex, bounded, closed subset $\mathcal{U}_{\mathsf{ad}}$ of a Hilbert space $\mathcal{U}$. For any $\mu \in \mathcal{P}_{\mathsf{ad}}$ the bilinear form $a_\mu : V \times V \to \mathbb{R}$ is

⋆ The authors would like to acknowledge the financial support within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG).

- symmetric:
$$a_\mu(\varphi,\phi) = a_\mu(\phi,\varphi) \quad \text{for all } \varphi,\phi \in V;$$
- continuous: there exists a constant $\overline{\gamma}_\mu > 0$ satisfying
$$|a_\mu(\varphi,\phi)| \le \overline{\gamma}_\mu \|\varphi\|_V \|\phi\|_V \quad \text{for all } \varphi,\phi \in V;$$
- coercive: there is $\underline{\alpha}_\mu > 0$ such that
$$a_\mu(\varphi,\varphi) \ge \underline{\alpha}_\mu \|\varphi\|_V^2 \quad \text{for all } \varphi \in V$$

Moreover, $f_\mu(\cdot\,;u) \in L^2(0;T;V')$ holds for any $(\mu,u) \in \mathcal{X}_{\mathsf{ad}} = \mathcal{P}_{\mathsf{ad}} \times \mathcal{U}_{\mathsf{ad}}$.

For any $\mu \in \mathcal{P}_{\mathsf{ad}}$ it follows that (1) admits a unique solution $y = y_\mu \in \mathcal{Y}$ with
$$\mathcal{Y} = W(0,T) = L^2(0,T;V) \cap H^1(0,T;V');$$
cf., e.g., Hinze et al. (2009); Tröltzsch (2010). Recall that
$$\tfrac{\mathrm{d}}{\mathrm{d}t} \langle \varphi(t), \phi \rangle_H = \langle \varphi_t(t), \phi \rangle_{V',V} \quad \text{for all } (\varphi,\phi) \in \mathcal{Y} \times V$$
and $\mathcal{Y} \hookrightarrow C([0,T];H)$. We assume that a_μ and f_μ depend affinely on the parameters:
$$a_\mu = \sum_{l=1}^{m_a} \vartheta_l^a(\mu) \hat{a}_l, \quad f_\mu(t;u) = \sum_{l=1}^{m_f} \vartheta_l^f(\mu) \hat{f}_l(t;u) \tag{2}$$
for any $(\mu,u) \in \mathcal{X}_{\mathsf{ad}}$ and $t \in [0,T]$. Otherwise, we apply the *(discrete) empirical interpolation method* to get approximations satisfying (2); cf., e.g., Barrault et al. (2004); Chaturantabut and Sorensen (2010, 2012); Hesthaven et al. (2016).

The goal is to estimate unknown model parameters $\mu \in \mathcal{P}_{\mathsf{ad}}$, where the control input should be chosen in an optimal way explained later. Here, we suppose that $u \in \mathcal{U}_{\mathsf{ad}}$ is fixed and consider
$$\min J(y,\mu) \text{ subject to } (y,\mu) \in \mathcal{Y} \times \mathcal{P}_{\mathsf{ad}} \text{ satisfies (1)} \quad (\mathbf{P})$$
As (1) is uniquely solvable, we can define the *reduced cost* $\hat{J}(\mu) = J(y_\mu,\mu)$, where y_μ solve (1). Then, $(\mathbf{P})$ is equivalent to
$$\min \hat{J}(\mu) \quad \text{subject to (s.t.)} \quad \mu \in \mathcal{P}_{\mathsf{ad}} \qquad (\hat{\mathbf{P}})$$

In our application the cost quantifies the discrepancy to a given desired (or observed) state $\hat{y} \in L^2(0,T;H)$:
$$\hat{J}(\mu) = \frac{1}{2} \int_0^T \|y_\mu(t) - \hat{y}(t)\|_H^2 \,\mathrm{d}t + \frac{\sigma}{2} \|\mu - \hat{\mu}\|_2^2 \tag{3}$$
where $\|\cdot\|_2$ stands for the Euclidean norm, σ is a non-negative weight and $\hat{\mu} \in \mathbb{R}^d$ is a reference parameter. Existence of an optimal solution $\bar{\mu}$ is guaranteed, but – due to non-convexity – there are possibly many local solutions; cf., e.g., Hinze et al. (2009). A local optimal solution $\bar{\mu}$ to $(\hat{\mathbf{P}})$ is characterized by first-order necessary optimality conditions. Let $\bar{y} = y_{\bar{\mu}}$ be the optimal state associated with $\bar{\mu}$ and the *adjoint variable* $\bar{p} = p_{\bar{\mu}} \in \mathcal{Y}$ the solution of the *adjoint equation*
$$\begin{aligned} -\tfrac{\mathrm{d}}{\mathrm{d}t} \langle \bar{p}(t), \varphi \rangle_H + a_{\bar{\mu}}(\bar{p}(t),\varphi) &= \langle \hat{y}(t) - \bar{y}(t), \varphi \rangle_H \\ \bar{p}(T) &= 0 \quad \text{in } H \end{aligned}$$
for all $\varphi \in V$ and $t \in [0,T)$. Using the *adjoint approach* the gradient of the cost functional at $\bar{\mu}$ is given as
$$\begin{aligned} \nabla \hat{J}(\bar{\mu}) = \sigma(\bar{\mu} - \hat{\mu}) &+ \textstyle\int_0^T \sum_{i=1}^{m_a} \nabla \vartheta_i^a(\mu) \hat{a}_q(\bar{y}(t),\bar{p}(t)) \,\mathrm{d}t \\ &- \textstyle\int_0^T \sum_{i=1}^{m_f} \nabla \vartheta_i^f(\mu) \langle \hat{f}_i(t;u), p(t) \rangle_{V',V} \,\mathrm{d}t \in \mathbb{R}^d. \end{aligned}$$

The derivation is based on the Lagrangian
$$\begin{aligned} \mathcal{L}(y,\mu,p;u) = J(y,\mu) &+ \int_0^T \langle y_t(t), p(t) \rangle_{V',V} \,\mathrm{d}t \\ &+ \int_0^T a_\mu(y(t),p(t)) - \langle f_\mu(t;u), p(t) \rangle_{V',V} \,\mathrm{d}t \end{aligned}$$
for $(\mathbf{P})$.

3. DISCRETIZATION

Next we introduce a high-dimensional discretization, called *full-order model* (FOM), which is accurate but expensive to solve. To reduce significantly the computational costs a further approximation is applied, the *reduced-order model* (ROM), faster to solve but less accurate.

3.1 FOM

Let $\varphi_1,\dots,\varphi_N \in V$ be given linearly independent functions and $V^N = \operatorname{span}\{\varphi_1,\dots,\varphi_N\} \subset V$. The FOM for (1) reads: for given $\mu \in \mathcal{P}_{\mathsf{ad}}$ the function $y_\mu^N(t) \in V^N$ solves
$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} \langle y_\mu^N(t), \varphi \rangle_H + a_\mu(y_\mu^N(t),\varphi) &= \langle f_\mu(t;u), \varphi \rangle_{V',V} \\ y_\mu^N(0) &= \mathcal{P}^N y_\circ \end{aligned} \tag{4}$$
for all $\varphi \in V^N$ and $t \in (0,T]$, where $\mathcal{P}^N : H \to V^N$ is a projection. Due to $y_\mu^N(t) \in V^N$ we have
$$y_\mu^N(t) = \sum_{i=1}^N \mathrm{y}_{\mu i}(t) \varphi_i \quad \text{for } t \in [0,T]$$
so that (4) reduces into finding the coefficient vector $\mathrm{y}_\mu(t) = (\mathrm{y}_{\mu i}(t))_{1 \le i \le N}$ solving
$$\begin{aligned} \mathrm{M}\dot{\mathrm{y}}_\mu(t) + \mathrm{A}_\mu \mathrm{y}_\mu(t) &= \mathrm{f}_\mu(t;u), \quad t \in (0,T] \\ \mathrm{M}\mathrm{y}_\mu(0) &= \mathrm{y}_\circ \end{aligned} \tag{5}$$
for $\mathrm{f}_\mu(t;u) = (\langle f_\mu(t;u), \varphi_i \rangle_{V';V}) \in \mathbb{R}^N$, $\mathrm{y}_\circ = (\langle y_\circ, \varphi_i \rangle_H) \in \mathbb{R}^N$, $\mathrm{A}_\mu = ((a_\mu(\varphi_j,\varphi_i))) \in \mathbb{R}^{N \times N}$, $\mathrm{M} = ((\langle \varphi_j, \varphi_i \rangle_H)) \in \mathbb{R}^{N \times N}$.

Remark 1. Due to (2) both A_μ and f_μ satisfy
$$\mathrm{A}_\mu = \sum_{l=1}^{m_a} \vartheta_l^a(\mu) \hat{\mathrm{A}}_l, \quad \mathrm{f}_\mu(t;u) = \sum_{l=1}^{m_f} \vartheta_l^f(\mu) \hat{\mathrm{f}}_l(t;u)$$
for $\hat{\mathrm{A}}_l = ((\hat{a}_l(\varphi_j,\varphi_i)))$ and $\hat{\mathrm{f}}_l(t;u) = (\langle \hat{f}_l(t;u), \varphi_i \rangle_{V';V})$. $\diamond$

For solving (5) we apply the implicit Euler method for the time integration; cf., e.g., Quarteroni (2017). To simplify the presentation we utilize an equidistant time grid $t_k = (k-1)\Delta t$, $k = 1,\dots,K$ and $\Delta t = T/(K-1)$. Then, the problem is to find $\{\mathrm{y}_\mu^k\}_{k=1}^K \subset \mathbb{R}^N$ solving
$$\begin{aligned} (\mathrm{M} + \Delta t \mathrm{A}_\mu)\mathrm{y}_\mu^k &= \mathrm{M}\mathrm{y}_\mu^{k-1} + \Delta t \mathrm{f}_\mu^k(u) \\ \mathrm{M}\mathrm{y}_\mu^1 &= \mathrm{y}_\circ \end{aligned} \tag{6}$$
for $k = 2,\dots,K$ and $\mu \in \mathcal{P}_{\mathsf{ad}}$ with $\mathrm{f}_\mu^k(u) = \mathrm{f}_\mu(t^k;u) \in \mathbb{R}^N$. Note that y_μ^k is the coefficient vector of the approximation $y_\mu^k \in V^N$ of $y_\mu^N(t_k) \in V^N$; cf. (10).

Remark 2. Similarly, we get the Galerkin approximations for the adjoint state, the reduced cost, and the reduced cost gradient. More precisely, the discrete adjoint sequence $\{\mathrm{p}_\mu^k\}_{k=1}^K \subset \mathbb{R}^N$ satisfies:
$$\begin{aligned} (\mathrm{M} + \Delta t \mathrm{A}_\mu)\mathrm{p}_\mu^k &= \mathrm{M}\big(\mathrm{p}_\mu^{k+1} + \Delta t(\hat{\mathrm{y}}^k - \mathrm{y}_\mu^k)\big) \\ \mathrm{p}_\mu^K &= 0 \end{aligned}$$

where $\{\mathrm{y}_\mu^k\}_{k=1}^K$ solves (6) and $\hat{y}^k = \sum_{i=1}^N \hat{\mathrm{y}}_i^k \varphi_i \approx \hat{y}(t_k)$ holds. Further, the reduced cost $\hat{J}$ is approximated by

$$\hat{J}^N(\mu) = \tfrac{1}{2}\sum_{k=1}^K \alpha_k \left(\mathrm{y}_\mu^k - \hat{\mathrm{y}}^k\right)^\top \mathrm{M}\left(\mathrm{y}_\mu^k - \hat{\mathrm{y}}^k\right) + \tfrac{\sigma}{2}\left\|\mu - \hat{\mu}\right\|_2^2 \quad (7)$$

where the α_k's are trapezoidal weights. The gradient of the reduced cost is then given by

$$\nabla \hat{J}^N(\mu) = \sum_{k=1}^K \alpha_k \Big[\sum_{l=1}^{m_a} \nabla \vartheta_l^a(\mu) \hat{\mathrm{A}}_l \mathrm{y}_\mu^k - \sum_{l=1}^{m_f} \nabla \vartheta_l^f(\mu) \hat{\mathrm{f}}_l^k(u)\Big]^\top \mathrm{p}_\mu^k + \sigma(\mu - \hat{\mu})$$

◇

3.2 ROM

The reduced order modelling is based on the projection of the model onto a smaller dimension space: in particular, we call the reduced-order space $V^\ell = \operatorname{span}\{\psi_i\}_{i=1}^\ell \subset V^N$ for $\ell \ll N$. We use the method of *proper orthogonal decomposition* (POD), that consists in finding such basis using the snapshots $\{\mathrm{y}_\mu^k\}_{k=1}^K \subset \mathbb{R}^N$, or equivalently $\{y_\mu^k\}_{k=1}^K \subset V^N$, satisfying the problem

$$\begin{aligned} &\min_{\{\psi_j\}_{j=1}^\ell} \sum_{k=1}^K \alpha_k \Big\| y_\mu^k - \sum_{i=1}^\ell \langle y_\mu^k, \psi_i\rangle_V \ \psi_i \Big\|_V^2 \\ &\text{subject to } \langle \psi_i, \psi_j\rangle_V = \delta_{ij} \qquad \forall\ 1 \le i,j \le \ell. \end{aligned} \quad (8)$$

A solution to (8) is given by the eigenvectors corresponding to the ℓ biggest eigenvalues of the matrix $R = \mathrm{YDY}^\top \mathrm{W}$, where $\mathrm{Y} = [\mathrm{y}_\mu^1, \dots, \mathrm{y}_\mu^K] \in \mathbb{R}^{N\times K}$ is the matrix of snapshots, D is the diagonal matrix of the trapezoidal weights and $\mathrm{W} = ((\langle \varphi_j, \varphi_i\rangle_V)) \in \mathbb{R}^{N\times N}$.

Remark 3. These ℓ eigenvectors can be easily evaluated by defining $\bar{\mathrm{Y}} = \mathrm{W}^{1/2}\mathrm{YD}^{1/2}$ and solving the eigenvalue problem

$$\begin{aligned} &\bar{\mathrm{Y}}^\top \bar{\mathrm{Y}} \mathrm{v}_i = \lambda_i \mathrm{v}_i, && 1 \le i \le \ell, \\ &\text{s.t. } \langle \mathrm{v}_i, \mathrm{v}_j\rangle_{\mathbb{R}^N} = \delta_{ij}, && 1 \le i,j \le \ell, \end{aligned}$$

where the matrix $\bar{\mathrm{Y}}^\top\bar{\mathrm{Y}} = \mathrm{D}^{1/2}\mathrm{YWYD}^{1/2}$ is relatively easy to evaluate (i.e. there is actually no need to evaluate the square root of W and the square root of a diagonal matrix is straightforward), and then defining the vectors

$$\Psi_i = \frac{1}{\sqrt{\lambda_i}} \mathrm{YD}^{1/2}\mathrm{v}_i \in \mathbb{R}^N$$

as the coefficient vectors of the reduced basis element ψ_i in the FOM basis $\{\varphi_i\}_{i=1}^N$, for $i = 1, \dots, \ell$. ◇

Furthermore, we have the following error formula:

$$\sum_{k=1}^K \alpha_k \Big\| y_\mu^k - \sum_{i=1}^\ell \langle y_\mu^k, \psi_i\rangle_V \ \psi_i \Big\|_V^2 = \sum_{j=\ell+1}^d \lambda_i. \quad (9)$$

We define the coefficient matrix $\Psi \in \mathbb{R}^{N\times\ell}$ satisfying $\psi_j = \sum_{i=1}^N \Psi_{ij}\varphi_i$ for $j = 1, \dots, \ell$. We notice how each column Ψ_i is the coefficient vector of ψ_i, coherently as defined before. Then, the ROM for (4) reads as follows: find $\{\mathrm{y}_\mu^{k,\ell}\}_{k=1}^K \subset \mathbb{R}^\ell$ such that

$$\begin{aligned} \left(\mathrm{M}^\ell + \Delta t \mathrm{A}_\mu^\ell\right)\mathrm{y}_\mu^{k,\ell} &= \mathrm{M}^\ell \mathrm{y}_\mu^{k-1,\ell} + \Delta t \mathrm{f}^{k,\ell}(u) \\ \mathrm{M}^\ell \mathrm{y}_\mu^{1,\ell} &= \mathrm{y}_\circ^\ell \end{aligned}$$

with $\mathrm{y}_\circ^\ell = \Psi^\top \mathrm{y}_\circ$, $\mathrm{M}^\ell = \Psi^\top \mathrm{M}\Psi$, $\mathrm{A}_\mu^\ell = \sum_{l=1}^{m_a} \vartheta_l^a(\mu)\Psi^\top \hat{\mathrm{A}}_l \Psi$ and $\mathrm{f}_\mu^{k,\ell}(u) = \sum_{l=1}^{m_f} \vartheta_l^f(\mu)\Psi^\top \hat{\mathrm{f}}_l^k(u)$. The solution $\{\mathrm{y}_\mu^{k,\ell}\}_{k=1}^K$ uniquely exists (see Kunisch and Volkwein (2001)) and it is interpreted as a reduced-order approximation for $\{\mathrm{y}_\mu^k\}_{k=1}^K$:

$$\mathrm{y}_\mu^k \approx \tilde{\mathrm{y}}_\mu^{k,\ell} := \Psi \mathrm{y}_\mu^{k,\ell} \in \mathbb{R}^N$$

For $1 \le k \le K$ we also use the additional notations

$$y_\mu^k = \sum_{i=1}^N \mathrm{y}_{\mu i}^k \varphi_i \in V^N \text{ and } y_\mu^{k,\ell} = \sum_{i=1}^\ell \mathrm{y}_{\mu i}^{k,\ell}\psi_i \in V^\ell \quad (10)$$

Similarly, we construct a reduced-order adjoint approximation $\{\mathrm{p}_\mu^{k,\ell}\}_{k=1}^K \subset \mathbb{R}^\ell$ by solving

$$\begin{aligned} \left(\mathrm{M}^\ell + \Delta t \mathrm{A}_\mu^\ell\right)\mathrm{p}_\mu^{k,\ell} &= \mathrm{M}^\ell \mathrm{p}_\mu^{k+1,\ell} + \Delta t\left(\hat{\mathrm{y}}^{k,\ell} - \mathrm{M}^\ell \mathrm{y}_\mu^{k,\ell}\right) \\ \mathrm{p}^{K,\ell} &= 0 \end{aligned}$$

with $\hat{\mathrm{y}}^{k,\ell} = \Psi^\top \mathrm{M}\hat{\mathrm{y}}^k$. Then, since

$$\begin{aligned} \|y_\mu^{k,\ell} - \hat{y}^k\|_H^2 = &\left(\mathrm{y}_\mu^{k,\ell}\right)^\top \mathrm{M}^\ell \mathrm{y}_\mu^{k,\ell} \\ &- 2\left(\mathrm{y}^{k,\ell}\right)^\top \hat{\mathrm{y}}^{k,\ell} + \left(\hat{\mathrm{y}}^k\right)^\top \mathrm{M}\hat{\mathrm{y}}^k, \end{aligned}$$

the reduced cost function is approximated by

$$\begin{aligned} \hat{J}^\ell(\mu) = &\tfrac{1}{2}\sum_{k=1}^K \alpha_k \left[\left(\mathrm{y}_\mu^{k,\ell}\right)^\top \mathrm{M}^\ell \mathrm{y}_\mu^{k,\ell} - 2\left(\mathrm{y}^{k,\ell}\right)^\top \hat{\mathrm{y}}^{k,\ell}\right] \\ &+ \tfrac{1}{2}\sum_{k=1}^K \alpha_k \left(\hat{\mathrm{y}}^k\right)^\top \mathrm{M}\hat{\mathrm{y}}^k + \tfrac{\sigma}{2}\left\|\mu - \hat{\mu}\right\|_2^2, \end{aligned} \quad (11)$$

where the element

$$\frac{1}{2}\sum_{k=1}^K \alpha_k \left(\hat{\mathrm{y}}^k\right)^\top \mathrm{M}\hat{\mathrm{y}}^k \approx \frac{1}{2}\|\hat{y}\|_{L^2(0,T;H)}^2$$

can be evaluated and stored as soon as the state $\hat{y}$ is evaluated. Finally, the gradient of the reduced cost function is given by

$$\begin{aligned} \nabla \hat{J}^\ell(\mu) = \sum_{k=1}^K \alpha_k \Big[&\sum_{l=1}^{m_a} \nabla \vartheta_l^a(\mu) \hat{\mathrm{A}}_l^\ell \mathrm{y}_\mu^{k,\ell} \\ &- \sum_{l=1}^{m_f} \nabla \vartheta_l^f(\mu) \hat{\mathrm{f}}_l^{k,\ell}(u)\Big]^\top \mathrm{p}_\mu^{k,\ell} + \sigma(\mu - \hat{\mu}). \end{aligned}$$

Let us use a variation of the standard POD: instead of finding the RB by solving the problem (8), we find it by solving

$$\begin{aligned} \min_{\{\psi_j\}_{j=1}^\ell} \Bigg\{ &\sum_{k=1}^K \alpha_k \Big\| y_\mu^k - \sum_{i=1}^\ell \langle y_\mu^k, \psi_i\rangle_V \ \psi_i \Big\|_V^2 \\ &+ \sum_{k=1}^K \alpha_k \Big\| p_\mu^k - \sum_{i=1}^\ell \langle p_\mu^k, \psi_i\rangle_V \ \psi_i \Big\|_V^2 \\ &+ \sum_{k=2}^K \Delta t \Big\| \bar{\partial} y^k - \sum_{i=1}^\ell \langle \bar{\partial} y^k, \psi_i\rangle_V \ \psi_i \Big\|_V^2 \\ &+ \sum_{k=2}^K \Delta t \Big\| \bar{\partial} p^k - \sum_{i=1}^\ell \langle \bar{\partial} p^k, \psi_i\rangle_V \ \psi_i \Big\|_V^2 \Bigg\} \\ \text{s.t.} &\langle \psi_\mathrm{i}, \psi_\mathrm{j}\rangle_\mathrm{V} = \delta_\mathrm{ij} \qquad \forall\ 1 \le \mathrm{i},\mathrm{j} \le \ell, \end{aligned}$$

Here, we have used the notation

$$\bar{\partial} y^k = \tfrac{y_\mu^k - y_\mu^{k-1}}{\Delta t} \approx (y_\mu)_t(t_k)$$

and

$$\bar{\partial} p^k = \tfrac{p_\mu^k - p_\mu^{k-1}}{\Delta t} \approx (p_\mu)_t(t_k)$$

for $k = 2, \dots, K$, to impose a certain level accuracy of the reduced basis also with respect to the adjoint variable and the approximations of the derivative of the states; cf. Gubisch and Volkwein (2017); Kunisch and Volkwein (2001). Indeed, defining

$$\begin{aligned} \mathrm{Z} = [&\mathrm{y}_\mu^1, \dots, \mathrm{y}_\mu^K, \mathrm{p}_\mu^1, \dots, \mathrm{p}_\mu^K \\ &\bar{\partial}\mathrm{y}_\mu^2, \dots, \bar{\partial}\mathrm{y}_\mu^K, \bar{\partial}\mathrm{p}_\mu^2, \dots, \bar{\partial}\mathrm{p}_\mu^K] \in \mathbb{R}^{N\times(4K-2)} \end{aligned} \quad (12)$$

and

$$\bar{\mathrm{D}} = \mathrm{diag}\,(\mathrm{D}, \mathrm{D}, \Delta t I_{K-1}, \Delta t I_{K-1})$$

where I is the identity matrix, we solve the eigenvalue problem

$$\begin{aligned} &\bar{\mathrm{Z}}^\top \bar{\mathrm{Z}} \tilde{\mathrm{v}}_i = \tilde{\lambda}_i \tilde{\mathrm{v}}_i, && 1 \le i \le \ell, \\ &\text{s.t. } \langle \tilde{\mathrm{v}}_i, \tilde{\mathrm{v}}_j \rangle_{\mathbb{R}^N} = \delta_{ij}, && 1 \le i, j \le \ell, \end{aligned} \tag{13}$$

where $\bar{\mathrm{Z}} = \mathrm{W}^{1/2} \mathrm{Z} \bar{\mathrm{D}}^{1/2}$. Hence, the error formula (9) becomes

$$\begin{aligned} &\sum_{k=1}^{K} \alpha_k \Big\| y_\mu^k - \sum_{i=1}^{\ell} \langle y_\mu^k, \psi_i \rangle_V \ \psi_i \Big\|_V^2 \\ &+ \sum_{k=1}^{K} \alpha_k \Big\| p_\mu^k - \sum_{i=1}^{\ell} \langle p_\mu^k, \psi_i \rangle_V \ \psi_i \Big\|_V^2 \\ &+ \sum_{k=2}^{K} \Delta t \Big\| \bar{\partial} y^k - \sum_{i=1}^{\ell} \langle \bar{\partial} y^k, \psi_i \rangle_V \ \psi_i \Big\|_V^2 \\ &+ \sum_{k=2}^{K} \Delta t \Big\| \bar{\partial} p^k - \sum_{i=1}^{\ell} \langle \bar{\partial} p^k, \psi_i \rangle_V \ \psi_i \Big\|_V^2 = \sum_{i=\ell+1}^{d} \tilde{\lambda}_i. \end{aligned}$$

For more details, see Kunisch and Volkwein (2001, 2002).

3.3 A-posteriori RB error estimates

Now we present a-posteriori error estimates for the error of the ROM without evaluating the full-order solution. For our TR algorithm we will need estimates for the state and for the cost function.

Let us define the μ-dependent space-time energy norm for the sequence $\{y^k\}_{k=1}^K \subset V^N$ as

$$\|y^j\|_\mu = \Big(\|y^j\|_H^2 + \Delta t \sum_{k=2}^{j} a_\mu(y^k, y^k) \Big)^{1/2}$$

for $j = 2, \dots, K$, and $\|y^1\|_\mu = \|y^1\|_H$. Then we can define the error estimates in this norm in the following propositions.

Proposition 4. Let $\{y_\mu^k\}_{k=1}^K$ and $\{y_\mu^{k,\ell}\}_{k=1}^K$ be given by (10) and $\mathsf{err}_\mu^k = y_\mu^k - y_\mu^{k,\ell} \in V^N$ for $k = 1, \dots, K$. For $\varphi \in V^N$ we define the residual

$$\begin{aligned} \langle \mathsf{res}_\mu^k, \varphi \rangle_{(V^N)', V^N} = &\Big\langle \tfrac{y_\mu^{k-1,\ell} - y_\mu^{k,\ell}}{\Delta t}, \varphi \Big\rangle_H - a_\mu(y_\mu^{k,\ell}, \varphi) \\ &+ \langle f_\mu(t_k; u), \varphi \rangle_{V',V}, \end{aligned}$$

and

$$\varepsilon_\mu^k = \|\mathsf{res}_\mu^k\|_{(V^N)'} = \sup_{\varphi \in V^N \setminus \{0\}} \frac{\langle \mathsf{res}_\mu^k, \varphi \rangle_{(V^N)', V^N}}{\|\varphi\|_V}$$

Then $\|\mathsf{err}_\mu^j\|_\mu \le \Delta_\mu^j$ for $j \in \{1, \dots, K\}$, where

$$\Delta_\mu^j = \left(\|\mathsf{err}_\mu^1\|_H^2 + \tfrac{\Delta t}{\underline{\alpha}_\mu} \sum_{k=2}^{j} \left| \varepsilon_\mu^j \right|^2 \right)^{1/2} \tag{14}$$

Proof. We refer to Grepl and Patera (2005). □

Proposition 5. Let $\{y_\mu^k\}_{k=1}^K$ and $\{y_\mu^{k,\ell}\}_{k=1}^K$ be given by (10), the reduced cost functions $\hat{J}^N(\mu)$ and $\hat{J}^\ell(\mu)$ be given by (7) and (11) respectively, the estimator Δ_μ^j be defined by (14), and the cost error defined by

$$\mathsf{err}_\mu^{\hat{J}} = \hat{J}^N(\mu) - \hat{J}^\ell(\mu).$$

Then we have

$$|\mathsf{err}_\mu^{\hat{J}}| \le \Delta_\mu^{\hat{J}}, \tag{15}$$

where

$$\Delta_\mu^{\hat{J}} = \tfrac{1}{2} \sum_{k=1}^{K} \alpha_k \big((\Delta_\mu^k)^2 + 2 \Delta_\mu^j \, \|y_\mu^{k,\ell} - \hat{y}^k\|_H \big), \tag{16}$$

which, again, does not require the evaluation of the full-order solution $\{y_\mu^k\}_{k=1}^K$.

Proof. First, we can write

$$\begin{aligned} &\|\hat{y}^k - y_\mu^k\|_H^2 - \|\hat{y}^k - y_\mu^{k,\ell}\|_H^2 \\ &= \langle \hat{y}^k - y_\mu^k, \hat{y}^k - y_\mu^k \rangle_H - \langle \hat{y}^k - y_\mu^{k,\ell}, \hat{y}^k - y_\mu^{k,\ell} \rangle_H \\ &= \langle \hat{y}^k - y_\mu^k, y_\mu^{k,\ell} - y_\mu^k \rangle_H - \langle y_\mu^k - y_\mu^{k,\ell}, \hat{y}^k - y_\mu^{k,\ell} \rangle_H \\ &= \langle \mathsf{err}_\mu^k, y_\mu^k + y_\mu^{k,\ell} - 2\hat{y}^k \rangle_H \\ &= \langle \mathsf{err}_\mu^k, \mathsf{err}_\mu^k + 2y_\mu^{k,\ell} - 2\hat{y}^k \rangle_H \\ &= \|\mathsf{err}_\mu^k\|_H^2 + 2 \langle \mathsf{err}_\mu^k, y_\mu^{k,\ell} - \hat{y}^k \rangle_H. \end{aligned}$$

Moreover, using the Cauchy-Schwarz inequality

$$\left| \langle \mathsf{err}_\mu^k, y_\mu^{k,\ell} - \hat{y}^k \rangle_H \right| \le \|\mathsf{err}_\mu^k\|_H \|y_\mu^{k,\ell} - \hat{y}^k\|_H$$

Then,

$$\begin{aligned} |\mathsf{err}_\mu^{\hat{J}}| &\le \tfrac{1}{2} \sum_{k=1}^{K} \alpha_k \left| \|\hat{y}^k - y_\mu^k\|_H^2 - \|\hat{y}^k - y_\mu^{k,\ell}\|_H^2 \right| \\ &\le \tfrac{1}{2} \sum_{k=1}^{K} \alpha_k \left| \|\mathsf{err}_\mu^k\|_H^2 + 2 \langle \mathsf{err}_\mu^k, y_\mu^{k,\ell} - \hat{y}^k \rangle_H \right| \\ &\le \tfrac{1}{2} \sum_{k=1}^{K} \alpha_k \big(\|\mathsf{err}_\mu^k\|_H^2 + 2 \|\mathsf{err}_\mu^k\|_H \|y_\mu^{k,\ell} - \hat{y}^k\|_H \big) \\ &\le \tfrac{1}{2} \sum_{k=1}^{K} \alpha_k \big((\Delta_\mu^k)^2 + 2 \Delta_\mu^j \, \|y_\mu^{k,\ell} - \hat{y}^k\|_H \big) \end{aligned}$$

Together with (16) we get (15).

4. TRUST-REGION RB APPROXIMATION

It is fairly easy to find a reduced basis for one fixed parameter, as we see in the next subsection. But when the parameter changes, the basis found before is not necessarily a good reduced basis for the new state. For this reason, a Greedy algorithm is often used (in the offline phase) to find a "common" reduced basis for all the parameters in the admissible set $\mathcal{P}_{\mathsf{ad}}$ so that in the online phase a fast computation of a state for a new admissible parameter is possible. This leads to the well-known offline/online decomposition (see, e.g., Grepl and Patera (2005), Binev et al. (2011), Haasdonk (2013)). In the context of optimization the offline computation might not be very suitable because during the optimization method only parameters from a small (but a-priorily unknown) subset of $\mathcal{P}_{\mathsf{ad}}$ are required. In Qian et al. (2017), Keil et al. (2021), Banholzer et al. (2020)) the RB space is built during a trust-region (TR) optimization process, where – following the optimization path – the RB space is enriched if it is necessary. We will utilize these ideas to develop our algorithm for the optimal experimental design. The full algorithm is presented in Algorithm 1, called for short *trust region reduced basis optimization* (TRRB-O), but for further details we refer to Banholzer et al. (2020) for further information. See also the recent paper Banholzer et al. (2022), where the TR approach is used in a multi-objective parameter optimization problem. Finally, we would like to refer to earlier work on methods based on TR proper orthogonal decomposition (TR-POD); see Arian et al. (2000); Sachs et al. (2014); Rogg et al. (2017).

Initialize the ROM at $\mu^{(0)}$, set $k=0$, $\delta^{(0)}=\delta_0$, *Loop_flag* $= True$.
while *Loop_flag* **do**
  BFGS sub-problem: find $\tilde{\mu}$ solution of (19) with stopping criteria (20). Store $\mu^{(k)}_{AGC}$ in the meantime.
  if (23) *is True* **then**
    Parameter is accepted: $\mu^{(k+1)}=\tilde{\mu}$.
    Set $\delta^{(k+1)}=\delta^{(k)}$.
    $\hat{J}^N$, $\nabla\hat{J}^N$, $\rho^{(k)}$, g^N, and $g^{\ell,(k)}$ are evaluated in $\mu^{(k+1)}$.
    if $g^h(\mu^{(k+1)})\leq\varepsilon_{tr}$ **then**
      Set *Loop_flag* $= False$.
    else
      if $\rho_k\geq\eta$ **then**
        Set $\delta^{(k+1)}=2\delta^{(k)}$.
      end
      if (26) *is False* **then**
        The model is enriched at $\mu^{(k+1)}$.
      end
    end
  else if (25) *if True* **then**
    The parameter $\tilde{\mu}$ is rejected.
    if *Skip_enrichment* $= False$ in the previous iteration **then**
      The model is enriched in $\tilde{\mu}$.
    end
    Set $\delta^{(k+1)}=\frac{1}{2}\delta^{(k)}/2$.
  else
    $\hat{J}^N$, $\nabla\hat{J}$, $\rho^{(k)}$, g^N, and $g^{\ell,(k)}$ are evaluated in $\tilde{\mu}$.
    if $g^N(\tilde{\mu})\leq\varepsilon_{tr}$ **then**
      Set $\mu^{(k+1)}=\tilde{\mu}$.
      Set *Loop_flag* $= False$.
    else
      if (26) *is True and* $\rho_k\geq\eta$ **then**
        The parameter is accepted: $\mu^{(k+1)}=\tilde{\mu}$.
        Set $\delta^{(k+1)}=2\delta^{(k)}$.
      else if $\hat{J}^N(\mu^{(k+1)})\leq\hat{J}^{\ell,(k)}(\mu^{(k)}_{ACG})$ **then**
        The parameter is accepted: $\mu^{(k+1)}=\tilde{\mu}$.
        The model is enriched in $\mu^{(k+1)}$.
        if $\rho_k\geq\eta$ **then**
          Set $\delta^{(k+1)}=2\delta^{(k)}$
        else
          Set $\delta^{(k+1)}=\delta^{(k)}$.
        end
      else
        The parameter $\tilde{\mu}$ is rejected.
        if *Skip_enrichment* $= False$ in the previous iteration **then**
          The model is enriched in $\tilde{\mu}$.
        end
        The TR radius is shrunken: $\delta^{(k+1)}=\delta^{(k)}/2$.
      end
    end
  end
  Set $k=k+1$.
end

Algorithm 1: Adaptive TR-RB optimization

4.1 Finding a reduced basis

For a fixed $\mu\in\mathcal{P}_{\mathsf{ad}}$ we compute a reduced-order approximation of y^k by the proper orthogonal decomposition (POD) method through the use of multiple snapshots, as explained previously. We find, then, a reduced basis $V^\ell=\text{span}\{\psi_1,\dots,\psi_\ell\}$ that can be characterized by the matrix Ψ,

$$\psi_i=\sum_{i=1}^{N}\Psi_{ij}\varphi_i\in V^N,\ \langle\psi_j,\psi_i\rangle_V=\delta_{ij},\ 1\leq i,j\leq\ell \quad (17)$$

The choice for the dimension of the reduced space ℓ may vary; for example, it can be fixed a-priorily or it can be chosen based on the eigenvalue decay in the POD method; see, e.g., Gubisch and Volkwein (2017). In our case, we decided to determine ℓ depending on the eigenvalues: given $\tilde{\lambda}_1\geq\tilde{\lambda}_2\geq\dots$ solutions of (13), we take the ℓ biggest eigenvalues such that $\lambda_{\ell+1}/\lambda_1<\varepsilon_{rel}$. This means that in the worst case possible $\sum_{i=\ell+1}^{K}\tilde{\lambda}_i<K\lambda_1\cdot\varepsilon_{rel}$; looking at our numerical examples, where $K=201$ and $\varepsilon_{rel}=10^{-6}$ or 10^{-7}, it is guaranteed that at least $\sum_{i=\ell+1}^{K}\tilde{\lambda}_i<\lambda_1\cdot 10^{-4}$. In our numerical experiment we have also fixed the maximum initial basis dimension to 10; we have observed that this does not impact on the overall accuracy.

4.2 Enriching a reduced basis

Assume that we already have computed an RB space characterized by the matrix Ψ (see (17)). Let us suppose that we want to enrich the basis in the parameter $\mu_+\in\mathcal{P}_{\mathsf{ad}}$. Then, the enrichment consists of finding another matrix $\Psi_+\in\mathbb{R}^{N\times\ell_+}$ using POD on the subspace that is orthogonal to the reduced space, namely substituting Z in (12) with $\tilde{\mathrm{Z}}$, whose columns are defined by

$$\tilde{\mathrm{Z}}_k=\mathrm{Z}_k-\sum_{i=1}^{\ell}\langle\mathrm{Z}_k,\Psi_i\rangle_{\mathrm{W}}\ \Psi_i.$$

Once we found the matrix Ψ_+, we merge it with Ψ, using Gram-Schmidt numerical W-orthonormalization, if necessary.

4.3 TR framework

The TR optimization method computes iteratively a first-order critical point of ($\hat{\mathbf{P}}$). At each iteration $k\geq 0$ of the optimization algorithm, we call the parameter $\mu^{(k)}$. We consider a cheaply computable model $m^{(k)}$ (approximation of the reduced cost) that can be trusted to accurately represent the function $\hat{J}$ in a reasonable neighborhood of $\mu^{(k)}$, called *trust region* $\mathcal{T}(\delta^{(k)})=\{\mu:\|\mu-\mu^{(k)}\|_2\leq\delta^{(k)}\}$, where $\delta^{(k)}$ is called *TR radius*. The TR method finds $\mu^{(k+1)}$ by solving the problem

$$\min_{s\in\mathbb{R}^d} m^{(k)}(s)\quad\text{s.t.}\quad\|s\|_2\leq\delta^{(k)},\ \mu^{(k)}+s\in\mathcal{P}_{\mathsf{ad}} \quad (18)$$

Setting $\tilde{\mu}=\mu^{(k)}+s$ the RB version of (18) is

$$\min_{\tilde{\mu}\in\mathcal{P}_{\mathsf{ad}}}\hat{J}^{\ell,(k)}(\tilde{\mu})\quad\text{s.t.}\quad q^{(k)}(\tilde{\mu})=\frac{\Delta^{\hat{J},(k)}_{\tilde{\mu}}}{\hat{J}^{\ell,(k)}(\tilde{\mu})}\leq\delta^{(k)} \quad (19)$$

Here and whenever some quantity depends on the iteration k, we show it in the superscript, like the RB cost $\hat{J}^{\ell,(k)}$.The

so-called efficiency $q^{(k)}$ helps us to quantify the accuracy of the RB and to define the TR.

How the model behaves in the TR tells us if we need to enrich the basis, or reduce the TR radius (in case the model is not accurate enough), or if we can enlarge the RB basis and even skip the enrichment process, in case the model is already accurate enough in the trust region.

4.4 TR subproblem

Let us suppose that we have fixed a TR radius $\delta^{(k)}$. Then a proposed parameter, solution of (19), is evaluated using a projected Armijo-BFGS algorithm (we refer to Kelley (1999) for the details). The iterations of the BFGS algorithm are indicated in the second subscript of the parameter: $\{\mu^{(k,j)}\}_{j=1}^{m_k}$ is the sequence of BFGS iterates and $\tilde{\mu} = \mu^{(k,m_k)}$ is the result of the BFGS algorithm. The maximum number of BFGS iterates is fixed at 400, and the algorithm finishes automatically when one of the termination criterion is satisfied:

$$q^{(k)}(\mu^{(k,j)}) \geq 0.95 \cdot \delta^{(k)} \tag{20a}$$

$$\|\mu^{(k,j)} - \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu^{(k,j)} + \nabla \hat{J}^{\ell,(k)}(\mu^{(k,j)}))\|_2 \leq 10^{-8} \tag{20b}$$

where (20a) tells us if we are too close to the border of the trust region (where the RB model is less accurate) and $\mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}$ is the projection onto $\mathcal{P}_{\mathsf{ad}}$.

Once that the projected BFGS method evaluates the descent direction $d^{(k,j)}$ for the parameter $\mu^{(k,j)}$, the Armijo backtracking finds $\mu^{(k,j+1)}$ as:

$$\mu^{(k,j+1)} = \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu^{(k,j)} + \alpha_{(k,j)} d^{(k,j)}) \in \mathcal{P}_{\mathsf{ad}}$$

where $\alpha_{(k,j)} = 0.5^{\beta}$ and the power $\beta = \beta(k,j)$ is the smallest integer such that the sufficient decrease and the TR constraint are satisfied:

$$\begin{aligned} \hat{J}^{\ell,(k)}(\mu^{(k,j+1)}) - \hat{J}^{\ell,(k)}(\mu^{(k,j)}) \\ \leq -\tfrac{\alpha_\circ}{\alpha_{(k,j)}} \|\mu^{(k,j+1)} - \mu^{(k,j)}\|_2^2 \end{aligned} \tag{21a}$$

with $\alpha_\circ = 10^{-4}$ and

$$q^{(k)}(\mu^{(k,j+1)}) \leq \delta^{(k)} \tag{21b}$$

4.5 Modification of the trust region

Once that the sub-problem has found a suitable candidate $\tilde{\mu} := \mu^{(k,m_k)}$, there are different possibilities: for example, we might realize that the candidate is on the border of the trust region, and since the model is not accurate there, we could shrink the TR radius and find a new candidate or just enrich the basis; on the other hand, if we notice that some "predicted sufficient reduction" is satisfied, we can enlarge the TR radius.

The approximated generalized Cauchy point (AGC) is defined as $\mu_{AGC}^{(k)} = \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu^{(k)} - \alpha^{(k,0)} \nabla \hat{J}^{\ell,(k)}(\mu^{(k)}))$, with $\alpha^{(k,0)}$ chosen such that (21a) and (21b) are satisfied. The cost of the evaluation of this parameter is "free", since it is evaluated in the first iteration of the BFGS algorithm. Then, an error-aware sufficient decrease condition (EASDC) is introduced (Qian et al. (2017)):

$$\hat{J}^{\ell,(k+1)}(\mu^{(k+1)}) \leq \hat{J}^{\ell,(k)}\big(\mu_{AGC}^{(k)}\big) \tag{22}$$

where we highlight that $\hat{J}^{\ell,(k)}$ refers to the reduced model at iteration k, while $\hat{J}^{\ell,(k+1)}$ refers instead to the model after the $(k+1)$-th (eventual) enrichment. This condition is central in the proof of convergence and the "sufficient" and "necessary" conditions we will see in the algorithm refer to the verification of it. Let us say explicitly that this condition is not straightforward to verify. Indeed, the left-hand side requires the evaluation of the cost function post-enhancement on the new parameter. We want to avoid this evaluation if some other conditions are not satisfied. For this reason, we postpone the evaluation of the FOM solution until we have to check the termination criterion or we want to enrich the RB.

To ensure that the candidate is a good parameter, a sufficient condition is analyzed:

$$\hat{J}^{\ell,(k)}(\tilde{\mu}) + \Delta_{\tilde{\mu}}^{\hat{J},(k)} < \hat{J}^{\ell,(k)}(\mu_{AGC}^{(k)}). \tag{23}$$

If (23) is satisfied, the candidate is accepted, $\mu^{(k+1)} = \tilde{\mu}$ and the model is "updated" (that means, the basis is enriched) there. Then, the TR radius is doubled if the predicted sufficient reduction (of model $m^{(k)}$) is realized, namely if

$$\rho^{(k)} = \tfrac{\hat{J}^N(\mu^{(k)}) - \hat{J}^N(\mu^{(k+1)})}{\hat{J}^{\ell,(k)}(\mu^{(k)}) - \hat{J}^{\ell,(k)}(\mu^{(k+1)})} \geq \eta = 0.75 \tag{24}$$

If the sufficient condition (23) does not hold, we check a necessary condition:

$$\hat{J}^{\ell,(k)}(\tilde{\mu}) - \Delta_{\tilde{\mu}}^{\hat{J},(k)} < \hat{J}^{\ell,(k)}\big(\mu_{AGC}^{(k)}\big) \tag{25}$$

If this also fails, it means that the point $\tilde{\mu}$ probably needs an enhancement too big to satisfy the error-aware sufficient decrease condition (22). Then it must be rejected, the model must be enriched and the TR radius shrunken. If, on the other hand, the sufficient condition fails and the necessary condition holds, we enrich the model and check (22) on the candidate parameter:

$$\hat{J}^{\ell,(k+1)}(\tilde{\mu}) \leq \hat{J}^{\ell,(k)}\big(\mu_{AGC}^{(k)}\big)$$

If this holds, the candidate and the enrichment are accepted (and again, if (24) holds the radius is doubled), while if it fails both the parameter and the enrichment are rejected and the radius is shrunken.

Defining

$$g^N(\mu^{(k+1)}) = \|\mu^{(k+1)} - \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu^{(k+1)} - \nabla \hat{J}^N(\mu^{(k+1)}))\|_2$$

the stopping criterion for the parameter estimation is

$$g^N(\mu^{(k+1)}) \leq \varepsilon_{tr}$$

where ε_{tr} is the overall tolerance, set as 1e-5.

4.6 Skipping the enrichment

Enriching the basis often leads to reduced bases with too many elements, hence wasting the reduced-order model purpose. For this reason, in Banholzer et al. (2020) the possibility of skipping the enrichment was included. In particular, this happens when all these three conditions are true:

$$q^{(k)}(\mu^{(k+1)}) \leq \tfrac{\delta^{(k+1)}}{2} \tag{26a}$$

$$\tfrac{|g^N(\mu^{(k+1)}) - g^{\ell,(k)}(\mu^{(k+1)})}{g^{\ell,(k)}(\mu^{(k+1)})} \leq 10^{-1} \tag{26b}$$

$$\tfrac{\|\nabla \hat{J}^N(\mu^{(k+1)}) - \nabla \hat{J}^{\ell,(k)}(\mu^{(k+1)})\|_2}{\|\nabla \hat{J}^N(\mu^{(k+1)})\|_2} \leq \min\left\{10^{-2}, \tfrac{\delta^{(k+1)}}{2}\right\} \tag{26c}$$

with $g^{\ell,(k)}(\mu) = \|\mu - \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu - \nabla \hat{J}^{\ell,(k)}(\mu))\|_2$. Inequality (26a) indicates how much the model $m^{(k+1)}$ is trustworthy,

inequality (26b) checks the convergence criterion of the reduced model, and inequality (26c) checks the RB accuracy of the cost gradient.

4.7 Some comments on the convergence of the algorithm

The convergence of the algorithm to a first-order critical point of $\hat{J}^N$ is showed in Keil et al. (2021) and Banholzer et al. (2020), Theorem 3.8, for the elliptic case. In our case the algorithm is the same, so we just have to make sure that the assumptions required for the convergence are satisfied. First of all, Assumption I of parameter separability of a_μ and f_μ is satisfied in (2). Then, we can assume a_μ, f_μ and J to be twice continuously Fréchet differentiable w.r.t. μ, as in Assumption II, which requires all the continuous functions θ_l^a and θ_l^f to be regular enough for $l = 1, \dots, m_a$ and $l = 1, \dots, m_f$, respectively. Assumption III simply requires the FOM to be accurate enough so that we can neglect its error. Assumption IV requires the cost functional $J(y, \mu)$ to be strictly positive for all $y \in \mathcal{Y}$ and $\mu \in \mathcal{P}_{\mathsf{ad}}$. Since at the moment we cannot guarantee this, we can add a constant to (3) without loss of generality, since in such a way we do not modify the minimization problem. Hence, we can define the new reduced cost function as

$$\hat{J}(\mu) = \frac{1}{2} \int_0^T \|y_\mu(t) - \hat{y}(t)\|_H^2 \, \mathrm{d}t + \frac{\sigma}{2} \|\mu - \hat{\mu}\|_2^2 + 1,$$

and redefine $\hat{J}^N$ and $\hat{J}^\ell$ similarly. Assumption V is a technical safe guard that we do not need to discuss more than it is in the original paper. Assumption VI requires the ROM gradient to be uniformly Lipschitz-continuous and its second derivatives to be locally Lipschitz-continuous, guaranteed in our case by the quadratic functional (11). We also assume the function $q^{(k)}$ to be uniformly continuous in $\mathcal{P}_{\mathsf{ad}}$ uniformly in k as in Assumption VII, which is in our case satisfied by the continuity of $q^{(k)}$ and the compactness of $\mathcal{P}_{\mathsf{ad}}$. Finally, let us make a comment on Lemma 3.4 of Banholzer et al. (2020). In the elliptic case, every time the model is enriched the following conditions are satisfied:

$$q^{(k+1)}(\mu^{(k+1)}) = 0, \quad \frac{\left| g^{\ell,(k+1)}(\mu^{(k+1)}) - g^N(\mu^{(k+1)}) \right|}{g^{\ell,(k)}(\mu^{(k+1)})} = 0,$$

but in our case this is not the case. For us it is sufficient, though, to require

$$q^{(k+1)}(\mu^{(k+1)}) \leq \delta^{(k+1)}/2$$

and

$$\frac{\left| g^{\ell,(k+1)}(\mu^{(k+1)}) - g^N(\mu^{(k+1)}) \right|}{g^{\ell,(k)}(\mu^{(k+1)})} \leq 10^{-1}$$

every time the model is enriched (i.e. two of the necessary conditions to skip the enrichment process) to satisfy Lemma 3.4.

5. OPTIMAL DESIGN OF EXPERIMENT

In the parameter estimation we have fixed the control u. In the current situation, we assume that we can run experiments for any control u and get an observed state $\hat{y}_u$, where the underlying true parameter is called μ^*. For each experiment we can find the optimal parameter $\bar{\mu}$ that minimizes the cost but $\bar{\mu} = \bar{\mu}_u$ depends on the choice of the control u. Simulating measurement errors on the observed state $\hat{y}_u$ we assume the data is a random variable presenting some stochastic noise. Hence, the solution $\bar{\mu} \in \mathcal{P}_{\mathsf{ad}}$ of the parameter estimation will be an outcome of a random parameter $\tilde{\mu} = \tilde{\mu}(u)$, whose expected value $\mathbb{E}(\tilde{\mu})$ and covariance matrix $\mathrm{C}(\tilde{\mu})$ also depends on the control u. The *optimal experimental design* (OED) deals with finding the control u that minimizes a certain measure $\phi(u)$ of uncertainty.

In particular, following what in literature is called A-criterion (Alexanderian et al. (2016)), we will minimize the trace of the covariance matrix; this is equivalent to minimizing the average mean square error of the maximum a-posteriori probability estimator or it can be seen as the minimization of the average half-axis length of the confidence ellipsoid with respect to the average variance of the parameters. Then, the optimal design $\bar{u}$ satisfies

$$\bar{u} = \arg\min \left\{ \phi(u) \,|\, u \in \mathcal{U}_{\mathsf{ad}} \right\} \quad \text{for } \phi(u) = \mathrm{trace}\left(\mathrm{C}(\tilde{\mu})\right)$$

In our discretized setting, the control $u = u(t)$ is a step function and it is determined by the vector $\mathrm{u} = (\mathrm{u}_i) \in \mathbb{R}^K$ with $u(t_k) = u_k$ for $k \in \{1, \dots, K\}$.

To get a least square problem, we start with a slight change of notation: the state now also depends on the control u and the cost $\hat{J}^N$ has the form

$$\hat{J}^N(\mu) = \tfrac{1}{2} \|\mathrm{F}(\mu, \mathrm{u})\|_2^2$$

where $\mathrm{F}(\mu, u) = \mathrm{h}(\mu, u) - \mathrm{b} \in \mathbb{R}^{NK+d+1}$ and

$$\mathrm{h}(\mu, u) = \begin{pmatrix} \left(\sqrt{\alpha_k} \mathrm{M}^{1/2} \mathrm{y}^k_{\mu,u}\right)_{1 \leq k \leq K} \\ \left(\sqrt{\sigma} \mu_i\right)_{1 \leq i \leq d} \\ 0 \end{pmatrix},$$

$$\mathrm{b} = \begin{pmatrix} \left(\sqrt{\alpha_k} \mathrm{M}^{1/2} \hat{\mathrm{y}}^k_u\right)_{1 \leq k \leq K} \\ \left(\sqrt{\sigma} \hat{\mu}_i\right)_{1 \leq i \leq d} \\ \sqrt{2} \end{pmatrix}.$$

We assume that the data has some Gaussian noise,

$$\hat{\mathrm{y}}^k_{u,i} = \mathrm{y}^k_{\mu^*,u,i} + \eta^k_i, \tag{27}$$

where $\eta^k_i \sim \mathcal{N}(0, \sigma_{\mathrm{b}}^2)$ for all $k = 1, \dots, K$ and $i = 1, \dots, N$. Then, the data vector is a random variable $\mathrm{b} \sim \mathcal{N}(\hat{\mathrm{b}}, \mathrm{C}^{\mathrm{b}})$, where

$$\hat{\mathrm{b}} = \begin{pmatrix} \left(\sqrt{\alpha_k} \mathrm{M}^{1/2} \mathrm{y}^k_{\mu^*,u}\right)_{1 \leq k \leq K} \\ \left(\sqrt{\sigma} \hat{\mu}_i\right)_{1 \leq i \leq d} \\ \sqrt{2} \end{pmatrix}$$

and C^{b} is a block-diagonal matrix, with

$$C^{\mathrm{b}} = \sigma_{\mathrm{b}}^2 \begin{pmatrix} \alpha_1 \mathrm{M} & & & & \\ & \ddots & & & \\ & & \alpha_K \mathrm{M} & & \\ & & & 0_d & \\ & & & & 0 \end{pmatrix},$$

where $0_d \in \mathbb{R}^{d \times d}$ contains only zeros.

The mean value of the estimator $\tilde{\mu}$ is the solution of the error-free parameter estimation, that is

$$m(\tilde{\mu}) = \arg\min_{\mu} \tfrac{1}{2} \sum_{k=1}^K \alpha_k \left\| y^k_{\mu^*,u} - \hat{y}^k_u \right\|_H^2 + \tfrac{\sigma}{2} \|\mu - \hat{\mu}\|^2 + 1$$

and regarding the covariance matrix we follow Bock et al. (2013): in the first-order approximation in the optimal pa-

rameter $\bar{\mu}$, the random parameter $\tilde{\mu}$ is distributed normally with a covariance matrix equal to

$$\mathrm{C}(\tilde{\mu};\mathrm{u}) = \mathrm{K}\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})^\top \mathrm{C}^\mathrm{b}\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})\mathrm{K}$$

where $\mathrm{K} = (\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})^\top \mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u}))^{-1}$ holds and $\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})$ is the Jacobian matrix of $\mathrm{F}(\bar{\mu};\mathrm{u})$, given by

$$\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u}) = (\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})_1 | \mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})_2),$$

where

$$\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})_1 = \begin{pmatrix} \left(\sqrt{\alpha_k}\mathrm{M}^{1/2}\mathrm{v}^{1,k}_{\bar{\mu},u}\right)_{1\le k\le K} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{NK+d+1}$$

and

$$\mathrm{J}_\mathrm{F}(\bar{\mu};\mathrm{u})_2 = \begin{pmatrix} \left(\sqrt{\alpha_k}\mathrm{M}^{1/2}\mathrm{v}^{2,k}_{\bar{\mu},u}\right)_{1\le k\le K} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{NK+d+1}$$

with $\{\mathrm{v}^{1,k}_{\bar{\mu},u}\}_{k=1}^K, \{\mathrm{v}^{2,k}_{\bar{\mu},u}\}_{k=1}^K \subset \mathbb{R}^N$ being the snapshots of the sensitivities of the state $y_{\bar{\mu},u}$. w.r.t. the first and the second components of the parameters, respectively.

Remark 6. Let us observe that we never need to evaluate the square root of the matrix M, since we only need the products $K = \mathrm{J}_\mathrm{F}^\top \mathrm{J}_\mathrm{F}$ and $J_\mathrm{F}^\top \mathrm{C}^d J_\mathrm{F}$.

In the numerical examples the OED optimization problem

$$\bar{\mathrm{u}} = \arg\min\left\{\phi(\mathrm{u}) \,|\, \mathrm{u} \in \mathrm{U}_{\mathsf{ad}}\right\}$$
$$\mathrm{U}_{\mathsf{ad}} = \left\{u \in \mathbb{R}^K \mid -3 \le \mathrm{u}_k \le 3 \text{ for } k = 1,\dots,K\right\}$$

is solved using the function `fmin_l_bfgs_b` from the Python library `scipy.optimize` without specifying the gradient of $\phi(u)$, approximated numerically. Also, to speed up the process, the time interval is divided into five sub-intervals and the optimization is done in each sub-interval.

6. NUMERICAL EXAMPLES

Our OED iterative algorithm is a bi-level cycle: starting from an initial admissible parameter and an initial control, the inner cycle is the TRRB-O, that finds the parameter optimizing the cost with the current control, while the outer cycle finds the control minimizing the uncertainty of such parameter. The overall algorithm stops when the ℓ_2-norm of two consecutive controls is less than 10^{-5}.

The initial reference parameter $\hat{\mu}$ in (3) is the middle point of $\mathcal{P}_{\mathsf{ad}}$, and at any following iteration the reference parameter is the optimal parameter found in the previous iteration. The constant σ is fixed to 10^{-8}.

We consider examples in a one-dimensional spatial interval, $\Omega = (0,1)$, with $H = L^2(\Omega)$ and $V = H^1(\Omega)$. In both cases the initial value is $y_\circ(x) \equiv 1$ and the time horizon is $T = 2$. The chosen initial control is $u(t) = \frac{1}{2}\cos(10t)/2$, and data is always evaluated with the formula (27), where $\mathrm{y}^k_{\mu^*,u}$ is evaluated with a FOM solver and $\sigma_b^2 = 10^{-4}$.

In the next tables we are going to analyze the number ℓ of bases generated by the TR optimization method, the error $\mathsf{err}^* = \|\bar{\mu} - \mu^*\|_2$ and $\phi(\bar{u}) = \min_u \phi(u)$, where the optimal solutions $\bar{\mu}$ and $\bar{u}$ are computed by the strategy explained in Section 5.

6.1 Two parameters

We consider $d = 2$ parameters, $\mu \in \mathcal{P}_{\mathsf{ad}} = [0.1, 1.7] \times [1, 3]$ and an equation in strong form

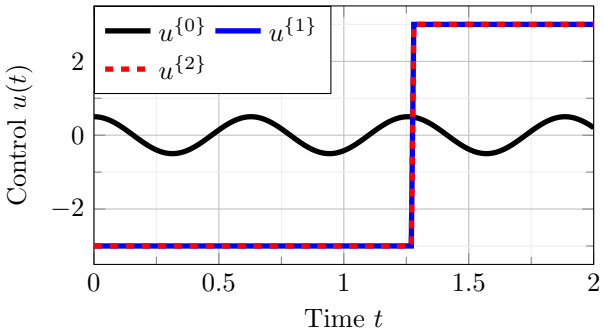
$$\begin{cases} y_t(x,t) - \mu_1 y_{xx}(x,t) + \frac{1}{2}y(x,t) = 0 \\ y_x(0,t) = 0, \qquad \mu_1 y_x(1,t) = \mu_2 u(t) \\ y(x,0) = y_\circ(x) \end{cases}$$

so that $a_\mu(y,\varphi) = \mu_1 \int_\Omega y'(x)\varphi'(x)\,\mathrm{d}x + \frac{1}{2}\int_\Omega y(x)\varphi(x)\mathrm{d}x$ and $\langle f_\mu(t;u),\varphi\rangle_{V',V} = \mu_2 u(t)\varphi(L)$. The true parameter is $\mu^* = (1,2)$ and the initial parameter is $\mu^{(0)} = (1.5, 1.5)$.

Table 1. Numerical example of Section 6.1.

	Iteration 1	Iteration 2
ℓ	10	7
μ_{opt}	$\begin{pmatrix} 1.00035 \\ 2.00023 \end{pmatrix}$	$\begin{pmatrix} 0.99996 \\ 1.99998 \end{pmatrix}$
err^*	4.20e-4	4.10e-5
TRRB-O time	11 *s*	5 *s*
$\phi(\bar{\mathrm{u}})$	1.13e-7	1.13e-7
OED time	99 *s*	4 *s*

Numerical example of Section 6.1.



First of all, we can observe how the TRRB-O algorithm converges very fast with only one initial basis formation. Also, the OED phase reaches the "best" control in just one iteration (blue). With this control the TRRB-O method finds a slightly better parameter and the OED finds in the same control (red, covering blue), therefore stopping.

6.2 Four parameters

Now $d = 4$ and $\mathcal{P}_{\mathsf{ad}} = [0.1, 1.7] \times [1.0, 3.0] \times [0.013, 4] \times [0.97, 2.22]$, and the problem is defined by

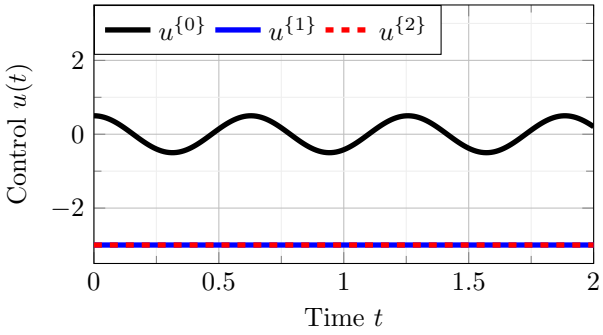
$$a_\mu(y,\varphi) = \mu_1 \textstyle\int_0^{0.2} y'(x)\varphi'(x)\,\mathrm{d}x + \mu_2 \int_{0.2}^{0.7} y'(x)\varphi'(x)\,\mathrm{d}x + \mu_3 \int_{0.7}^{1} y'(x)\varphi'(x)\mathrm{d}x + \frac{1}{2}\int_\Omega y(x)\varphi(x)\,\mathrm{d}x$$

and $\langle f_\mu(t;u),\varphi\rangle_{V',V} = \mu_4 u(t)\varphi(1)$. The real parameter is $\mu^* = (1, 1.3, 0.8, 2)$ and the initial parameter is $(1.5, 1, 1.2, 1.5)$. The result are showed in the next table.

Table 2. Numerical example of Section 6.2.

	Iteration 1	Iteration 2
ℓ	17	17
$\bar{\mu}$	$\begin{pmatrix} 1.4914 \\ 1.2803 \\ 0.8030 \\ 2.0003 \end{pmatrix}$	$\begin{pmatrix} 0.9996 \\ 1.3000 \\ 0.8000 \\ 2.0000 \end{pmatrix}$
err^*	0.49	3.9e-4
TRRB-O time	38 s	36 s
$\phi(\bar{u})$	3.9e-4	8.3e-5
OED time	882 s	9 s

Numerical example of Section 6.2.



In this example we can see how computational times and errors rise with bigger sizes of the parameter space. Let us observe that the TRRB-O algorithm needs 17 bases and works way better than a standard weak greedy and we can appreciate the full potential of such algorithm when applied to higher dimensions. Regarding the optimal design, as before the algorithm finds the optimal control in one iteration (blue and red overlap).

7. CONCLUSION

In this work we have considered an algorithm for the optimal design of experiment of a parameter-dependent evolution problem. Constructing the algorithm on a bi-level optimization framework, the idea of the creation and the enrichment of a basis on-the-go is preferred to an offline-online computation.

Starting from the parabolic problems in our numerical examples, extensions can include nonlinearities in the parameters, where for example it is not trivial evaluating a-posteriori error estimates for the ROM. In the same way, the a-posteriori error estimates we use are strongly based on the coercivity of the bilinear form, hence it would be of interest analyzing non-coercive problems.

Another possible extension is the addition of an algebraic state in the evolution system, which is typical, for example, in lithium-ion battery models. In that case the differential state (i.e. the ion concentration) is not observable and measurements are possible only on the algebraic state (the potential).

REFERENCES

Alexanderian, A., Gloor, P.J., and Ghattas, O. (2016). On Bayesian A- and D-optimal experimental designs in infinite dimensions. *Bayesian Anal.*, 11(3), 671–695.

Arian, E., Fahl, M., and Sachs, E. (2000). Trust-region proper orthogonal decomposition for flow control. Technical Report 2000-25, ICASE.

Banholzer, S., Keil, T., Mechelli, L., Ohlberger, M., Schindler, F., and Volkwein, S. (2020). An adaptive projected Newton non-conforming dual approach for trust-region reduced basis approximation of PDE-constrained parameter optimization. arXiv:2012.11653.

Banholzer, S., Mechelli, L., and Volkwein, S. (2022). A trust region reduced basis Pascoletti-Serafini algorithm for multi-objective PDE-constrained parameter optimization. arXiv:2201.07744.

Barrault, M., Maday, Y., Nguyen, N., and Patera, A. (2004). An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9), 667–672.

Benner, P., Grivet-Talocia, S., Quarteroni, A., Rozza, G., Schilders, W., and Silveira, L. (2020a). *Model Order Reduction: Volume 1: Basic Concepts and Notation*. De Gruyter.

Benner, P., Grivet-Talocia, S., Quarteroni, A., Rozza, G., Schilders, W., and Silveira, L. (2020b). *Model Order Reduction: Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter.

Benner, P., Grivet-Talocia, S., Quarteroni, A., Rozza, G., Schilders, W., and Silveira, L. (2020c). *Model Order Reduction: Volume 3: Applications*. De Gruyter.

Benner, P., Sachs, E., and Volkwein, S. (2014). Model order reduction for PDE constrained optimization. In *Trends in PDE constrained optimization*, volume 165 of *Internat. Ser. Numer. Math.*, 303–326. Birkhäuser/Springer, Cham.

Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., and Wojtaszczyk, P. (2011). Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.*, 43(3), 1457–1472.

Bock, H.G., Körkel, S., and Schlöder, J.P. (2013). Parameter estimation and optimum experimental design for differential equation models. In *Model based parameter estimation*, volume 4 of *Contrib. Math. Comput. Sci.*, 1–30. Springer, Heidelberg.

Chaturantabut, S. and Sorensen, D. (2010). Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5), 2737–2764.

Chaturantabut, S. and Sorensen, D. (2012). A state space estimate for POD-DEIM nonlinear model reduction. *SIAM Journal on Numerical Analysis*, 50(1), 46–63.

Grepl, M.A. and Patera, A.T. (2005). A posteriori error bounds for reduced-bias approximations of parametrized parabolic partial differential equations. *M2AN Math. Model. Numer. Anal.*, 39(1), 157–181.

Gubisch, M. and Volkwein, S. (2017). Proper orthogonal decomposition for linear-quadratic optimal control. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox (eds.), *Model Reduction and Approximation: Theory and Algorithms*, 5–66. SIAM, Philadelphia, PA.

Haasdonk, B. (2013). Convergence rates of the POD-greedy method. *ESAIM Math. Model. Numer. Anal.*, 47(3), 859–873.

Hesthaven, J., Rozza, G., and Stamm, B. (2016). *Certified Reduced Basis Methods for Parametrized Partial*

Differential Equations. SpringerBriefs in Mathematics. Springer Cham.

Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S. (2009). *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York.

Keil, T., Mechelli, L., Ohlberger, M., Schindler, F., and Volkwein, S. (2021). A non-conforming dual approach for adaptive trust-region reduced basis approximation of PDE-constrained parameter optimization. *ESAIM Math. Model. Numer. Anal.*, 55(3), 1239–1269.

Kelley, C.T. (1999). *Iterative methods for optimization*, volume 18 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, PA.

Körkel, S., Kostina, E., Bock, H., and Schlöder, J.P. (2004). Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optim. Method. Softw.*, 19, 327–338.

Kunisch, K. and Volkwein, S. (2001). Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90(1), 117–148.

Kunisch, K. and Volkwein, S. (2002). Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis*, 40, 492–515.

Pozzi, A., Ciaramella, G., Volkwein, S., and Raimondo, D. (2019). Optimal design of experiments for a lithium-ion cell: parameter identification of a single particle model with electrolyte dynamics. *Industrial and Engineering Chemistry Research*, 58, 1286–1299.

Qian, E., Grepl, M., Veroy, K., and Willcox, K. (2017). A certified trust region reduced basis approach to PDE-constrained optimization. *SIAM J. Sci. Comput.*, 39(5), S434–S460.

Quarteroni, A. (2017). *Numerical models for differential problems*, volume 16 of *MS&A. Modeling, Simulation and Applications*. Springer, Cham.

Reddy, S.R., Scharrer, M.K., Pichler, F., Watzenig, D., and Dulikravich, G.S. (2019). Accelerating parameter estimation in Doyle-Fuller-Newman model for lithium-ion batteries. *COMPEL*, 38, 1533–1544.

Rogg, S., Trenz, S., and Volkwein, S. (2017). Trust-region POD using a-posteriori error estimation for semilinear parabolic optimal control problems. Technical report, University of Konstanz. Konstanzer Schriften in Mathematik No. 359.

Sachs, E., Schneider, M., and Schu, M. (2014). Adaptive trust-region POD methods in PIDE-constrained optimization. *Numerische Mathematik*, 165, 327–342.

Tröltzsch, F. (2010). *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society.