Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-26 (2022) 149–154

# Trust-Region RB Methods for PDE-Constrained Optimization and Optimal Input Design

**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 input design involving a parameter-dependent evolution problem. 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.

Copyright © 2022 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

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

## 1. INTRODUCTION

Mathematical models based on partial differential equations (PDEs) are very useful in many fields, such as natural sciences, medicine and engineering. However, PDE models involve unknown parameters which have to be estimated from experiments, while we might have some control input variables that we are able to change in real time. Parameter estimation (PE) comprehends methods and algorithms are 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 PE, hence being an important choice when the temporal or computational costs are limited.

In this work two problems are discussed. First, the step of PE 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 methods are developed, where the "expensive" models are replaced by cheaper and less accurate surrogates. 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 our taylored optimization method in the reduced space is very close to a local minimum in the full space.

⋆ 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).

The second problem belongs to the optimal input/experimental design; we want to find the "best" control, namely the one that gives us a "better" PE in a bi-level algorithm. 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.

## 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 $t \in (0, T]$, 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 $\mathcal{U}$. Further, the bilinear form $a_\mu : V \times V \to \mathbb{R}$ is continuous, coercive, symmetric and $f_\mu(\cdot\,; u) \in L^2(0; T; V')$ for any $(\mu, u) \in \mathcal{X}_{\mathsf{ad}} = \mathcal{P}_{\mathsf{ad}} \times \mathcal{U}_{\mathsf{ad}}$.

It follows that (1) admits a unique solution $y = y_\mu \in \mathcal{Y} = W(0,T) = L^2(0,T;V) \cap H^1(0,T;V')$ for any $\mu \in \mathcal{P}_{\mathsf{ad}}$; cf., e.g., Hinze et al. (2009). Furthermore, we assume that $a_\mu$ and $f_\mu$ 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 *empirical interpolation method* to get approximations satisfying (2); cf., e.g., Hesthaven et al. (2016).

2405-8963 Copyright © 2022 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2022.10.392

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1vbfrcg1wrvu6

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 $y_\mu$ solves (1). Then, (**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 $\hat{y} \in L^2(0,T;H)$:

$$\hat{J}(\mu) = 1 + \tfrac{1}{2}\int_0^T \|y_\mu(t) - \hat{y}(t)\|_H^2 \,\mathrm{d}t + \tfrac{\sigma}{2}\|\mu - \hat{\mu}\|_2^2 \qquad (3)$$

where $\|\cdot\|_2$ stands for the Euclidean norm, $\sigma$ is a non-negative weight and $\hat{\mu} \in \mathbb{R}^d$ is a reference parameter. Existence of an optimal solution $\bar{\mu}$ can be ensured under continuity for $\mu \to a_\mu$ and $\mu \to f_\mu$, 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

$$\begin{aligned} \nabla\hat{J}(\bar{\mu}) = \sigma(\bar{\mu} - \hat{\mu}) &+ \int_0^T \sum_{i=1}^{m_a} \nabla\vartheta_i^a(\mu)\hat{a}_q(\bar{y}(t),\bar{p}(t))\,\mathrm{d}t \\ &- \int_0^T \sum_{i=1}^{m_f} \nabla\vartheta_i^f(\mu)\langle \hat{f}_i(t;u), p(t)\rangle_{V',V}\,\mathrm{d}t \in \mathbb{R}^d \end{aligned}$$

## 3. DISCRETIZATION

Next we introduce a high-dimensional discretization, called *full-order model* (FOM), which we assume to be accurate enough, but however 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

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\langle y_\mu^N(t), \varphi\rangle_H + a_\mu(y_\mu^N(t),\varphi) &= \langle f_\mu(t;u), \varphi\rangle_{V',V} \\ y_\mu^N(0) &= \mathcal{P}^N y_\circ \end{aligned} \qquad (4)$$

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

$$\begin{aligned} \mathrm{M}\dot{\mathrm{y}}_\mu(t) + \mathrm{A}_\mu \mathrm{y}_\mu(t) &= \mathrm{f}_\mu(t;u), \quad t \in (0,T] \\ \mathrm{M}\mathrm{y}_\mu(0) &= \mathrm{y}_\circ \end{aligned} \qquad (5)$$

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 $\mathrm{A}_\mu$ and $\mathrm{f}_\mu$ 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 $\hat{\mathrm{f}}_l(t;u) = (\langle \hat{f}_l(t;u), \varphi_i\rangle_{V';V})$. ◇

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

$$\begin{aligned} (\mathrm{M} + \Delta t\mathrm{A}_\mu)\mathrm{y}_\mu^k &= \mathrm{M}\mathrm{y}_\mu^{k-1} + \Delta t\mathrm{f}_\mu^k(u) \\ \mathrm{M}\mathrm{y}_\mu^1 &= \mathrm{y}_\circ \end{aligned} \qquad (6)$$

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 $\mathrm{y}_\mu^k$ is the coefficient vector of the approximation $y_\mu^k \in V^N$ of $y_\mu^N(t_k) \in V^N$; cf. (8).

*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 $\hat{J}$ in (3) is approximated by

$$\begin{aligned} \hat{J}^N(\mu) = 1 + \tfrac{1}{2}\sum_{k=1}^K \alpha_k \big(\mathrm{y}_\mu^k - \hat{\mathrm{y}}^k\big)^\top \mathrm{M}\big(\mathrm{y}_\mu^k - \hat{\mathrm{y}}^k\big) \\ + \tfrac{\sigma}{2}\|\mu - \hat{\mu}\|_2^2 \end{aligned} \qquad (7)$$

where the $\alpha_k$'s are trapezoidal weights. Then,

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

is the gradient of the reduced cost. ◇

### 3.2 ROM

For the moment we assume that for $\ell \ll N$ we are given an $\ell$-dimensional reduced-order space $V^\ell = \operatorname{span}\{\psi_i\}_{i=1}^\ell \subset V^N$, whose construction will be explained in Section 4. We define the coefficient matrix $\Psi \in \mathbb{R}^{N\times\ell}$ satisfying $\psi_j = \sum_{i=1}^N \Psi_{ij}\varphi_i$ for $j = 1, \dots, \ell$. Then, the ROM for (4) reads as follows: find $\{\mathrm{y}_\mu^{k,\ell}\}_{k=1}^K \subset \mathbb{R}^\ell$ such that

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

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

$$y_\mu^k = \sum_{i=1}^N \mathrm{y}_{\mu i}^k\varphi_i \in V^N \text{ and } y_\mu^{k,\ell} = \sum_{i=1}^\ell \mathrm{y}_{\mu i}^{k,\ell}\psi_i \in V^\ell \qquad (8)$$

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

$$\big(\mathrm{M}^\ell + \Delta t \mathrm{A}_\mu^\ell\big)\mathrm{p}_\mu^{k,\ell} = \mathrm{M}^\ell \mathrm{p}_\mu^{k+1,\ell} + \Delta t\big(\hat{\mathrm{y}}^{k,\ell} - \mathrm{M}^\ell \mathrm{y}_\mu^{k,\ell}\big)$$
$$\mathrm{p}^{K,\ell} = 0$$

Finally, the reduced cost function is approximated by

$$\begin{aligned}\hat{J}^\ell(\mu) = 1 + \tfrac{1}{2}\sum_{k=1}^K \alpha_k \left[\big(\mathrm{y}_\mu^{k,\ell}\big)^\top \mathrm{M}^\ell \mathrm{y}_\mu^{k,\ell} - 2\big(\mathrm{y}^{k,\ell}\big)^\top \hat{\mathrm{y}}^{k,\ell}\right] \\ + \tfrac{1}{2}\sum_{k=1}^K \alpha_k \big(\hat{\mathrm{y}}^k\big)^\top \mathrm{M}\hat{\mathrm{y}}^k + \tfrac{\sigma}{2}\left\|\mu - \hat{\mu}\right\|_2^2\end{aligned} \tag{9}$$

and its gradient is defined similarly.

### 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 trust-region (TR) algorithm we will need estimates for the state and for the cost function.

Let us define the $\mu$-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 $\|y^1\|_\mu = \|y^1\|_H$. Furthermore, let $\underline{\alpha}_\mu > 0$ be the $\mu$-dependent coercitivity constant of the bilinear form satisfying $a_\mu(\varphi, \varphi) \geq \underline{\alpha}_\mu \|\varphi\|_V^2$ for all $\varphi \in V$ and $\mu \in \mathcal{P}_{\mathsf{ad}}$. Then we can define the error estimates in the following prepositions.

*Proposition 3.* Let $\{y_\mu^k\}_{k=1}^K$ and $\{y_\mu^{k,\ell}\}_{k=1}^K$ be given by (8) 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 $\varepsilon_\mu^k = \|\mathsf{res}_\mu^k\|_{(V^N)'}$. Then $\|\mathsf{err}_\mu^j\|_\mu \leq \Delta_\mu^j$ for $j \in \{1, \dots, K\}$, where

$$\Delta_\mu^j = \Big(\|\mathsf{err}_\mu^1\|_H^2 + \tfrac{\Delta t}{\underline{\alpha}_\mu}\sum_{k=2}^j \big|\varepsilon_\mu^k\big|^2\Big)^{1/2} \tag{10}$$

**Proof.** We refer to Grepl and Patera (2005) for a proof.

*Proposition 4.* Let $\{y_\mu^k\}_{k=1}^K$ and $\{y_\mu^{k,\ell}\}_{k=1}^K$ be given by (8), the reduced cost functions $\hat{J}^N(\mu)$ and $\hat{J}^\ell(\mu)$ be given by (7) and (9) respectively, the estimator $\Delta_\mu^j$ be defined by (10), and the cost error defined by $\mathsf{err}_\mu^{\hat{J}} = \hat{J}^N(\mu) - \hat{J}^\ell(\mu)$. Then, we have

$$|\mathsf{err}_\mu^{\hat{J}}| \leq \Delta_\mu^{\hat{J}} \tag{11}$$

where

$$\Delta_\mu^{\hat{J}} = \tfrac{1}{2}\sum_{k=1}^K \alpha_k\big((\Delta_\mu^k)^2 + 2\Delta_\mu^k \|y_\mu^{k,\ell} - \hat{y}^k\|_H\big), \tag{12}$$

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

**Proof.** We skip the proof, available in Petrocchi et al. (2022).

## 4. TR-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 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. Since we cannot recall all details of the algorithm we refer to Petrocchi et al. (2022) and 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.

### 4.1 Creating and enriching 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 (cf., e.g., Kunisch and Volkwein (2001)). Given some relevant snapshots $\{\mathrm{z}^k\}_{k=1}^K \subset \mathbb{R}^N$, we use POD to compute a coefficient matrix $\Psi \in \mathbb{R}^{N\times\ell}$ by solving

$$\min\left\{\sum_{k=1}^K \alpha_k \|\mathrm{z}_\mu^k - \Psi\Psi^\top \mathrm{W}\mathrm{z}_\mu^k\|_{\mathrm{W}}^2 \;\middle|\; \Psi^\top \mathrm{W}\Psi = I\right\}$$

with the positive, symmetric matrix $\mathrm{W} = ((\langle\varphi_j, \varphi_i\rangle_V))$. It turns out that we have

$$\psi_i = \sum_{i=1}^N \Psi_{ij}\varphi_i \in V^N,\ \langle\psi_j, \psi_i\rangle_V = \delta_{ij},\ 1 \leq i, j \leq \ell, \tag{13}$$

so the reduced space can be characterized by the matrix $\Psi$. In general, the dimension of the reduced space $\ell$ 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). Details about our choice for the RB dimension and about our choice of snapshots are discussed in Petrocchi et al. (2022).

If, on the other hand, we already have computed an RB space $V^\ell = \operatorname{span}\{\psi_1, \dots, \psi_\ell\}$ characterized by the matrix $\Psi$ (see (13)), let us assume that we want to enrich the basis in the parameter $\mu_+ \in \mathcal{P}_{\mathsf{ad}}$. Then, we can find another matrix $\Psi_+ \in \mathbb{R}^{N\times\ell_+}$ using POD on the subspace that is orthogonal to the reduced space, and then we merge the matrices (using Gram-Schmidt W-orthonormalization, if necessary). Again, details are available in Petrocchi et al. (2022).

### 4.2 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 such approximated optimal 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

$$\min_{s\in\mathbb{R}^d} m^{(k)}(s) \quad \text{s.t.} \quad \|s\|_2 \leq \delta^{(k)}, \mu^{(k)} + s \in \mathcal{P}_{\mathsf{ad}} \tag{14}$$

Setting $\tilde{\mu} = \mu^{(k)} + s$ the RB version of (14) is

$$\min_{\tilde{\mu}\in\mathcal{P}_{\mathsf{ad}}} \hat{J}^{\ell,(k)}(\tilde{\mu}) \quad \text{s.t.} \quad q^{(k)}(\tilde{\mu}) = \frac{\Delta^{\hat{J},(k)}_{\tilde{\mu}}}{\hat{J}^{\ell,(k)}(\tilde{\mu})} \leq \delta^{(k)} \tag{15}$$

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.

### *4.3 TR subproblem*

Let us suppose that we have fixed a TR radius $\delta^{(k)}$. Then a proposed parameter, solution of (15), 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 superscript 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:

$$q^{(k)}(\mu^{(k,j)}) \geq \beta_1 \delta^{(k)} \tag{16a}$$

$$\|\mu^{(k,j)} - \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu^{(k,j)} + \nabla \hat{J}^{\ell,(k)}(\mu^{(k,j)}))\|_2 \leq \varepsilon_{sub}, \tag{16b}$$

where $\beta_1 \in (0, 1]$ and $\varepsilon_{sub}$ is the tolerance of the sub-problem. Equation (16a) 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:

$$\hat{J}^{\ell,(k)}(\mu^{(k,j+1)}) - \hat{J}^{\ell,(k)}(\mu^{(k,j)}) \leq -\frac{\alpha_\circ}{\alpha_{(k,j)}} \|\mu^{(k,j+1)} - \mu^{(k,j)}\|_2^2 \tag{17a}$$

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

$$q^{(k)}(\mu^{(k,j+1)}) \leq \delta^{(k)} \tag{17b}$$

### *4.4 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 $\mu^{(k)}_{AGC} = \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu^{(k)} - \alpha^{(k,0)} \nabla \hat{J}^{\ell,(k)}(\mu^{(k)}))$, with $\alpha^{(k,0)}$ chosen such that (17a) and (17b) are satisfied; cf. Yue and Meerbergen (2013). 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 is introduced (Qian et al. (2017)):

$$\hat{J}^{\ell,(k+1)}(\mu^{(k+1)}) \leq \hat{J}^{\ell,(k)}\big(\mu^{(k)}_{AGC}\big) \tag{18}$$

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:

$$\hat{J}^{\ell,(k)}(\tilde{\mu}) + \Delta^{\hat{J},(k)}_{\tilde{\mu}} < \hat{J}^{\ell,(k)}(\mu^{(k)}_{AGC}). \tag{19}$$

If (19) 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

$$\rho^{(k)} = \frac{\hat{J}^N(\mu^{(k)}) - \hat{J}^N(\mu^{(k+1)})}{\hat{J}^{\ell,(k)}(\mu^{(k)}) - \hat{J}^{\ell,(k)}(\mu^{(k+1)})} \geq \beta_2, \tag{20}$$

where $\beta_2 \in (0, 1)$.

If the sufficient condition (19) does not hold, we check a necessary condition:

$$\hat{J}^{\ell,(k)}(\tilde{\mu}) - \Delta^{\hat{J},(k)}_{\tilde{\mu}} < \hat{J}^{\ell,(k)}\big(\mu^{(k)}_{AGC}\big)$$

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 (18). 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 (18) on the candidate parameter:

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

If this holds, the candidate and the enrichment are accepted (and again, if (20) 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 $\varepsilon_{tr}$ is the overall tolerance.

### 4.5 Skipping the enrichment

Enriching the basis often leads to too many basis elements, hence wasting the ROM 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 hold:

$$q^{(k)}(\mu^{(k+1)}) \leq \beta_3 \delta^{(k+1)} \tag{21a}$$

$$\frac{|g^N(\mu^{(k+1)}) - g^{\ell,(k)}(\mu^{(k+1)})|}{g^{\ell,(k)}(\mu^{(k+1)})} \leq \tau_g \tag{21b}$$

$$\frac{\|\nabla \hat{J}^N(\mu^{(k+1)}) - \nabla \hat{J}^{\ell,(k)}(\mu^{(k+1)})\|_2}{\|\nabla \hat{J}^N(\mu^{(k+1)})\|_2} \leq \min\left\{\tau_{grad}, \beta_3 \delta^{(k+1)}\right\} \tag{21c}$$

with $\beta_3 \in (0,1)$, $g^{\ell,(k)}(\mu) = \|\mu - \mathcal{P}_{\mathcal{P}_{\mathsf{ad}}}(\mu - \nabla \hat{J}^{\ell,(k)}(\mu))\|_2$, and $\tau_g, \tau_{grad} > 0$ are arbitrary accuracies. Inequality (21a) indicates how much the model $m^{(k+1)}$ is trustworthy, inequality (21b) checks the convergence criterion of the reduced model, and inequality (21c) checks the RB accuracy of the cost gradient.

Convergence results for this algorithm are showed in Keil et al. (2021) and Banholzer et al. (2020) for an elliptic system. The proof can be adapted to our case, as discussed in Petrocchi et al. (2022).

## 5. OPTIMAL INPUT DESIGN

The result of the parameter estimation can depend strongly on the chosen control, even leading to convergence errors. In order to gain confidence in the accuracy of the parameter estimation, we want to find a "optimal" control, with the purpose of minimizing the uncertainty of the estimated parameter.

Following standard ideas of the optimal experimental/input design (Goodwin and Payne (1977), Atkinson et al. (2007)), we use an adaptive algorithm storing past data, such that the cost function will measure the discrepancy will all data stored. The superscript $[n]$ on objects like data vectors, state variables and cost functions specifies the experiment index, namely what control function it is being considered. The algorithm starts from initial guesses $u^{[1]}$ and $\mu^{[1]}$ and finds a new parameter and a new control at each iteration. Then, for $n > 1$, the algorithm iterates on three phases:

- The first phase is the simulation of some data $\hat{\mathrm{y}}^{[n-1]}$ using the control $u^{[n-1]}$ and the true parameter $\mu^*$.
- The second phase is the TR-RB approximation, from which we obtain the parameter $\mu^{[n]}$.
- The third phase is the optimal input design, whose purpose is to find a new control $u^{[n]}$. The algorithm stops when the control just found is very close to the previous one.

Simulating measurement errors, for $n \geq 1$ (experiment index) we assume that the data is a random variable presenting some additive noise, i.e.

$$\hat{\mathrm{y}}_i^{k,[n]} = \mathrm{y}_{\mu^* i}^{k,[n]} + \eta_{n,k,i}, \tag{22}$$

where $\eta_{n,k,i} \sim \mathcal{N}(0, \sigma_{\mathrm{d}}^2)$ for all $k = 1, \ldots, K$ (temporal index), $i = 1, \ldots, N$ (coefficient index in the FOM basis). So, in a way, we assume we are able to measure the FOM coefficient of the state variable corresponding to the true parameter at each time step, plus some error. The data collected at each iteration, $\hat{\mathrm{y}}^{[n]} \in \mathbb{R}^{NK}$, is a vector stacked with all entries $\hat{\mathrm{y}}_i^{k,[n]}$ for $k = 1, \ldots, K$ and $i = 1, \ldots, N$.

Hence, the solution of the second phase $\mu^{[n]}$ will be an estimator of the true parameter $\mu^*$. To evaluate the quality of such estimator we use the Fisher's information matrix, which quantifies how informative an experiment is.

Skipping the details, the Fisher's information matrix $\mathrm{M}_{\mu^{[n]}} \in \mathbb{R}^{d \times d}$ of the estimator $\mu^{[n]}$ is in our case

$$(\mathrm{M}_{\mu^{[n]}})_{j,l} = \frac{1}{\sigma_{\mathrm{d}}^2} \int_0^T \mathrm{s}_{\mu^{[n]},j}^{[n]}(t)^\top \mathrm{s}_{\mu^{[n]},l}^{[n]}(t) \mathrm{d}t$$

for any $j, l \in \{1, \ldots, d\}$, where $\mathrm{s}_{\mu^{[n]},j}^{[n]}(t)$ is the coefficient vector (in the FOM basis) of the sensitivity of the output with respect to the $j$-th parameter, evaluated in $\mu^{[n]}$ and using control $u^{[n]}$. More information about how this is evaluated (in particular in the RB framework) can be found in Petrocchi et al. (2022).

Then, the measure of uncertainty is chosen as the trace of the inverse of the Fisher's information matrix:

$$\phi(u) = \mathrm{trace}\,(\mathrm{M}_{\mu^{[n]}}^{-1})$$

## 6. NUMERICAL EXAMPLES

We consider examples in a one-dimensional spatial interval $\Omega = (0,4)$ 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 $u(t) = \frac{1}{2}\cos(10t) \in \mathcal{U}_{\mathsf{ad}} = \{u \in L^2(0,T) \,|\, -3 \leq u(t) \leq 3,\ \forall t \in (0,T)\} \subset \mathcal{U} = L^2(0,T)$, and data is always evaluated with the formula (22), where $\mathrm{y}_{\mu^*}^{k,[n]}$ is evaluated with a FOM solver and $\sigma_b^2 = 10^{-3}$. The initial reference parameter $\hat{\mu}$ in (3) is the middle point of $\mathcal{P}_{\mathsf{ad}}$, and at any following iteration $n > 1$ the reference parameter is the optimal parameter $\mu^{[n-1]}$ found in the previous iteration. The constant $\sigma$ is fixed to $10^{-8}$.

In the discretized setting, the control $u = u(t)$ is a step function and it is determined by the vector $\mathrm{u} = (\mathrm{u}_k) \in \mathbb{R}^K$ with $u(t_k) = u_k$ for $k \in \{1, \ldots, K\}$. Furthermore, in the numerical example the optimal input optimization problem

$$\mathrm{u}^{[n]} = \arg\min\left\{\phi(\mathrm{u}) \,|\, \mathrm{u} \in \mathrm{U}_{\mathsf{ad}}\right\}$$
$$\mathrm{U}_{\mathsf{ad}} = \left\{\mathrm{u} \in \mathbb{R}^K \,|\, -3 \leq \mathrm{u}_k \leq 3 \text{ for } k = 1, \ldots, K\right\}$$

is solved using the function `fmin_l_bfgs_b` from the Python library `scipy.optimize` without specifying the gradient of $\phi(u)$ and approximating it numerically.

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

**Run 1** (2 parameters). We consider $d = 2$ parameters, $\mu \in \mathcal{P}_{\mathsf{ad}} = [0.24, 3.4] \times [0.11, 5.0]$ 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 $a_\mu(y,\varphi) = \mu_1 \int_\Omega y'(x)\varphi'(x)\,\mathrm{d}x + \frac{1}{2}\int_\Omega y(x)\varphi(x)dx$ 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 $\mu^{[1]} = (3.4, 0.11)$.

Table 1. Run 1.

| | Starting values | $n=2$ | $n=3$ |
|---|---|---|---|
| $\ell_y, \ell_p$ | | 12, 30 | 18, 50 |
| err* | 3.05 | 3.14e-05 | 1.02e-06 |
| Second phase time | | 11 *s* | 9 *s* |
| $\phi(\mathrm{u}^{[n]})$ | 4.5e-06 | 1.8e-09 | 1.8e-09 |
| Third phase time | | 288 *s* | 31 *s* |

We can observe how the second phase converges very fast to a parameter very close to the true parameter. Furthermore, the third phase reaches the "best" control in just one iteration. With this control the second phase in the next iteration finds a more accurate parameter and then the third phase does not find a better control, therefore stopping.

**Run 2** (4 parameters) Let $d = 4$ and $\mathcal{P}_{\mathsf{ad}} = [0.24, 3.4] \times [0.11, 5] \times [0.013, 4] \times [0.97, 2.22]$ and

$$\begin{aligned} a_\mu(y,\varphi) = \mu_1 \textstyle\int_0^{1.5} y'(x)\varphi'(x)\,\mathrm{d}x + \mu_2 \int_{1.5}^{3} y'(x)\varphi'(x)\,\mathrm{d}x \\ + \mu_3 \textstyle\int_3^4 y'(x)\varphi'(x)dx + \frac{1}{2}\int_\Omega y(x)\varphi(x)\,\mathrm{d}x \end{aligned}$$

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 $(3, 5, 0.1, 2)$. The result are showed in the next table.

Table 2. Run 2.

| | Starting values | $n=2$ | $n=3$ |
|---|---|---|---|
| $\ell_y, \ell_p$ | | 24, 30 | 30, 40 |
| err* | 4.26 | 3.23 | 6.9e-05 |
| Second phase time | | 180 *s* | 20 *s* |
| $\phi(\mathrm{u}^{[n]})$ | 19 | 1.3e-06 | 2e-07 |
| Third phase time | | 540 *s* | 15 *s* |

In this example we can see how computational times and errors rise with bigger sizes of the parameter space. Let us observe that the TR algorithm works way faster than a standard weak greedy and we can appreciate the full potential of such algorithm when applied to higher dimensions. With the first control the algorithm fails to approximate the true parameter, but as soon as a new control is generated, the trust-region approximation gives a good estimator.

## REFERENCES


Atkinson, A.C., Donev, A.N., and Tobias, R.D. (2007). *Optimum experimental designs, with SAS*, volume 34 of *Oxford Statistical Science Series*. Oxford University Press, Oxford.

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. To appear in Pure and Applied Functional Analysis, issue 5, 2022.

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.

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.

Goodwin, G.C. and Payne, R.L. (1977). *Dynamic system identification*, volume 136 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London. Experiment design and data analysis.

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.

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

Petrocchi, A., Scharrer, M.K., and Volkwein, S. (2022). Adaptive reduced basis methods for PDE-constrained optimization and optimal input design. Technical report, Universität Konstanz, Konstanzer Schriften in Mathematik. URL `http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1l9b7uuht7b1s0`.

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.

Yue, Y. and Meerbergen, K. (2013). Accelerating optimization of parametric linear systems by model order reduction. *SIAM J. Optim.*, 23(2), 1344–1370.