A weighted POD-reduction approach for parametrized PDE-constrained optimal control problems with random inputs and applications to environmental sciences

A weighted POD-reduction approach for parametrized PDE-constrained Optimal Control Problems with random inputs and applications to environmental sciences

Giuseppe Carere*, Maria Strazzullo†, Francesco Ballarin‡,
Gianluigi Rozza† and Rob Stevenson*

October 20, 2021

Abstract

Reduced basis approximations of Optimal Control Problems (OCPs) governed by steady partial differential equations (PDEs) with random parametric inputs are analyzed and constructed. Such approximations are based on a Reduced Order Model, which in this work is constructed using the method of weighted Proper Orthogonal Decomposition. This Reduced Order Model then is used to efficiently compute the reduced basis approximation for any outcome of the random parameter. We demonstrate that such OCPs are well-posed by applying the adjoint approach, which also works in the presence of admissibility constraints and in the case of non linear-quadratic OCPs, and thus is more general than the conventional Lagrangian approach. We also show that a step in the construction of these Reduced Order Models, known as the aggregation step, is not fundamental and can in principle be skipped for noncoercive problems, leading to a cheaper online phase. Numerical applications in three scenarios from environmental science are considered, in which the governing PDE is steady and the control is distributed. Various parameter distributions are taken, and several implementations of the weighted Proper Orthogonal Decomposition are compared by choosing different quadrature rules.

1 Introduction

The search for a solution to a PDE-constrained optimization problem is in practice affected by unavoidable uncertainties, associated with measurements of parameters involved in the optimization problem. In such contexts of Uncertainty Quantification, one models these parameters as random variables. If the solution is a measurable function of the parameters, then it is a random variable itself, and one can study statistics that depend on the solution. For example, one may be interested in a moment of the solution or of a measurable function of the solution. In order to estimate such moments, a Monte Carlo type of estimator could be constructed, which takes the average over a large amount of solutions of the optimization problem, each corresponding to some outcome of the random parameter. When these solutions are not explicitly available, they can be approximated with accurate yet computationally expensive methods, such as Finite Element based methods. Consequently, the use of these so-called truth approximations in the construction of the estimator can lead to significant computational costs. To accommodate this issue, a Reduced Order Model (ROM) can be used to accelerate the approximation process, by providing a cheaply computable surrogate of the expensive truth approximation for any given parameter value, called a reduced basis approximation. The interested reader may refer to [Hesthaven et al., 2016, Prud'homme et al., 2001, Rozza et al., 2013, Rozza et al., 2008] for a survey on ROM techniques and to [Bader et al., 2017, Dedè, 2010, Kärcher et al., 2018, Negri et al., 2015, Negri et al., 2013] for their application to parametrized Optimal Control Problems, to [Torlo et al., 2018, Venturi et al., 2019a, Venturi et al., 2019b] for their application to UQ and to [Chen et al., 2017] for the application to OCPs in UQ.

*Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 105.

†mathLab, Mathematics Area, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, I-34136 Trieste, Italy.

‡Department of Mathematics and Physics, Catholic University of the Sacred Heart via Garzetta 48, I-25133 Brescia, Italy

Given a parametric measurement $\mu \in \mathbb{R}^n$ for $n \in \mathbb{N}$, in the general formulation of an Optimal Control Problem (OCP) parametrized by μ, one is to minimize a convex functional $J(\cdot, \cdot; \mu) : Y \times U \to \mathbb{R}$ over all state-control pairs $(y, u) \in Y \times U$ that satisfy the governing PDE-state equation $e(y, u; \mu) = 0$. Here, Y and U are the real Hilbert spaces of state and control, respectively, and $e(\cdot, \cdot; \mu) : Y \times U \to Q$ with Q some real Hilbert space. The minimizer $(y(\mu), u(\mu))$ is often constrained to lie in a closed and convex set of admissible pairs W_{ad} of $Y \times U$. Our focus, however, mainly lies on OCPs with $W_{ad} = Y \times U$, as it allows for the use of a ROM, and with $J(\cdot, \cdot; \mu)$ of quadratic form.
It is well-known (see e.g. [Lions, 1971]), that the minimizer $(y(\mu), u(\mu))$ can be obtained by finding the solution $(y(\mu), u(\mu), p(\mu)) \in Y \times U \times Q^*$ to some system of three equations, as shall also be recalled in this work. The variable $p(\mu)$ here is called the adjoint solution. To incorporate uncertainty in the parametric measurement μ, we interpret μ as the outcome of a random variable that takes values in some subset $\mathbb{M} \subset \mathbb{R}^n$. For any outcome $\mu \in \mathbb{M}$, the solution $(y(\mu), u(\mu), p(\mu))$ is in general not explicitly available, and one could construct the computationally expensive truth solution $(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu))$ to approximate it. As indicated above, the aim is to solve the OCP for a large number of possible outcomes of this random variable, so that, for example, sample averages can be computed. Not only would a naive parameter wise computation of the truth approximation become computationally infeasible, it would also completely ignore any possible low-dimensional behaviour of the (discrete) solution manifold

$$\mathcal{M}^{\mathcal{N}} := \left\{(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu)),\ \mu \in \mathbb{M}\right\} \subset Y \times U \times Q^*.$$

Indeed, the Kolmogorov N-width

$$\inf_{\substack{V \subset \operatorname{span}\{\mathcal{M}^{\mathcal{N}}\} \\ \dim V = N}} \operatorname*{ess\,sup}_{\mu \in \mathbb{M}} \|(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu)) - P_V(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu))\|_{Y \times U \times Q^*}$$

where P_V denotes the orthogonal projector onto V, very often decays rapidly in N. A ROM aims to exploit this by constructing a low dimensional subspace of $\operatorname{span}\{\mathcal{M}^{\mathcal{N}}\}$, in a computationally expensive so-called offline phase, onto which for any given outcome $\mu \in \mathbb{M}$ the solution $\left(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu)\right)$ then can be projected. This projection is performed in the online phase, which under certain parameter separability conditions can be executed rapidly, and results in a reduced basis approximation $(y_N(\mu), u_N(\mu), p_N(\mu))$ of the function $(y(\mu), u(\mu), p(\mu))$. The low dimensional subspace should be constructed in some optimal way. In this work, we do this by means of the weighted Proper Orthogonal Decomposition algorithm, already proposed in [Venturi et al., 2019a]. This algorithm is a combination of a singular value decomposition and a quadrature rule. See also [Ballarin et al., 2015, Burkardt et al., 2006, Chapelle et al., 2013].
Clearly, before one can construct truth approximations and reduced basis approximations of the control problem corresponding to an outcome $\mu \in \mathbb{M}$, it must be guaranteed that the OCP is well-posed for that μ. One way of doing this, is by using Brezzi's Theorem to search for saddle points of the Lagrangian corresponding to the OCP, as in [Negri et al., 2015, Negri et al., 2013, Strazzullo et al., 2017]. We shall present a second approach, also known as the adjoint approach or the Lions approach, based on the argumentation of [Lions, 1971, Chapters I, II] and [Hinze et al., 2009, Chapter I]. It allows one to recover the conclusions of the Lagrangian approach, but in a more general form, as it does not require $e(\cdot, \cdot; \mu)$ to be linear, $J(\cdot, \cdot; \mu)$ to be quadratic and W_{ad} to be $Y \times U$.
The work is outlined as follows. First of all, we formally define the formulation of our OCP in an Uncertainty Quantification context in Section 2. We then briefly recall the main arguments of the adjoint approach in Section 3, where the discussion is generalized from Hilbert to Banach spaces. For convenience, all linear spaces are assumed to be real.
The well-posedness of the truth approximations and reduced basis approximations is derived with similar argumentation as for the original OCP, as argued in Sections 4 and 5, respectively. For the reduced formulation, recent developments of ROMs for parametrized OCPs have made use of so-called aggregate spaces to ensure that the reduced formulation is well-posed, [Bader et al., 2017, Dedè, 2010, Kärcher et al., 2018, Negri et al., 2015, Negri et al., 2013]. We shall argue that this result can be improved by showing that the aggregation procedure is actually redundant from a theoretical point of view, and only useful in the specific case of coercive governing PDEs. This leads to an additional acceleration in the performance of the ROM, as the low dimensional space constructed in the offline phase can be taken even smaller. Section 5 also recalls the basic ideas of ROMs and the weighted Proper Orthogonal Decomposition.
Finally, in Section 6, we reconstruct the three applications of [Strazzullo et al., 2017] in marine sciences. Parametric uncertainties are inherent to such real world applications, and it therefore is important to

embed them in the Uncertainty Quantification context. We apply the weighted Proper Orthogonal Decomposition method with different choices of the quadrature rule to construct ROMs that accurately and efficiently approximate the OCP for arbitrary draws of the random parameter. We shall also consider several compactly supported probability distributions on $\mathbb{M}$ from which these draws are taken. The first numerical application involves a coercive PDE in weak formulation. The other two involve noncoercive (weakly coercive) PDEs, and for these we compare the performance of ROMs obtained with and without the aggregation procedure. Conclusions follow in Section 7.
The main novelty of the work lies in the derivation of well-posedness of ROMs for OCPs via the adjoint approach including their associated validity on Banach spaces, nonreflexive state spaces, and non-aggregated spaces, as well as the numerical application of these ROMs to existing OCP models embedded in a UQ context.

2 Problem formulation

Let $(\Theta, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbb{M}$ be a compact subset of $\mathbb{R}^n$ for $n \in \mathbb{N}$. Let also $\boldsymbol{\mu} : (\Theta, \mathcal{F}, \mathbb{P}) \to \mathbb{M}$ be a random variable with respect to the Borel σ-algebra on $\mathbb{M}$. We denote its law, i.e. the push forward measure of $\mathbb{P}$ under $\boldsymbol{\mu}$, by $\mathbb{P}^{\boldsymbol{\mu}}$. Furthermore, we let the Hilbert spaces Y and U be the *state space* and *control space* respectively and let $J(\cdot,\cdot;\cdot) : Y \times U \times \mathbb{M} \to \mathbb{R}$ denote the *objective functional*. Finally, let $e(\cdot,\cdot;\cdot) : Y \times U \times \mathbb{M} \to Q$ be the *state equation function* for some Hilbert space Q. We are interested in solving the following parametrized OCP:

Problem 2.1. Let us find $(y(\boldsymbol{\mu}), u(\boldsymbol{\mu})) : \Theta \to Y \times U$ such that it $\mathbb{P}$-almost everywhere holds that

- $e(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu}) = 0$,
- $(y(\boldsymbol{\mu}), u(\boldsymbol{\mu})) \in W_{ad} \subset Y \times U$,
- $J(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu}) = \min_{(\tilde{y},\tilde{u}) \in Y \times U} J(\tilde{y}, \tilde{u}; \boldsymbol{\mu})$

If $\mathbb{P}^{\boldsymbol{\mu}}$ corresponds to the uniform distribution on $\mathbb{M}$, then we say that the parametrized OCP is *deterministic*. We are primarily interested in the following particular kind of parametrized OCP. By $\mathcal{B}(H_1, H_2)$ we denote the space of bounded linear operators between the normed spaces H_1 and H_2.

Definition 2.2. Let Z be a Hilbert space, the so-called *observation space*, $z_d(\boldsymbol{\mu}) : \Theta \to Z$ a *desired* solution profile, $C \in \mathcal{B}(Y, Z)$ an *observation operator*. Define the quadratic objective functional $J(\cdot,\cdot;\boldsymbol{\mu})$

$$J(y, u; \boldsymbol{\mu}) = \frac{1}{2}\langle M(\boldsymbol{\mu})(Cy - z_d(\boldsymbol{\mu})), Cy - z_d(\boldsymbol{\mu})\rangle_{Z^* Z} + \frac{1}{2}\langle L(\boldsymbol{\mu})u, u\rangle_{U^* U}, \tag{2.1}$$

with $M(\boldsymbol{\mu}) : \Theta \to \mathcal{B}(Z, Z^*),\ L(\boldsymbol{\mu}) : \Theta \to \mathcal{B}(U, U^*)$ both self-adjoint.
Let $e(\cdot,\cdot,\boldsymbol{\mu})$ be affine, that is, there exist $A(\boldsymbol{\mu}) : \Theta \to \mathcal{B}(Y, Q),\ B(\boldsymbol{\mu}) : \Theta \to \mathcal{B}(U, Q),\ g(\boldsymbol{\mu}) : \Theta \to Q$ such that

$$e(y, u; \boldsymbol{\mu}) = A(\boldsymbol{\mu})y + B(\boldsymbol{\mu})u - g(\boldsymbol{\mu}) \quad \forall y \in Y, u \in U. \tag{2.2}$$

Then the corresponding Problem 2.1 is called a *linear-quadratic* OCP.

3 Solutions of Optimal Control Problems

In this section we describe the well-posedness of an OCP, temporarily dropping the parameter from the notation. The description is generalized from Hilbert to Banach spaces Y, U and Q. Throughout this section, we make the following assumption.

Assumption 3.1. It holds that

1. $W_{ad} = Y \times U_{ad}$ for $U_{ad} \subset U$,
2. for every $u \in U_{ad}$ there exists a unique $y = y(u) \in Y$ with $e(y, u) = 0$.

While the first assumption is made for the sake of simplicity, the second assumption is fundamental for the use of the adjoint approach, as it allows us to consider the following (unparametrized) reformulation of Problem 2.1:

Problem 3.2. Let us minimize $\tilde{J}(u) := J(y(u), u)$ over all $u \in U_{ad}$.

Existence and uniqueness of Problem 3.2 is discussed in the following proposition. If $\tilde{J}$ is G-differentiable (Gateaux differentiable) in u, then we denote the G-derivative in u by $D\tilde{J}(u) \in \mathcal{B}(U, U^*)$.

Proposition 3.3. *Let us suppose that Assumption 3.1 holds and that U_{ad} is nonempty and convex.*

(i) For the uniqueness of the solution of Problem 3.2, it suffices that $\tilde{J}$ is strictly convex.

(ii) For the existence of a solution of Problem 3.2, the following set of conditions is sufficient: U is reflexive, U_{ad} is closed, $\tilde{J}$ is weakly lower semicontinuous and $\tilde{J}(u) \to \infty$ whenever $\|u\|_U \to \infty$ in U_{ad}.

(iii) If $u \in U_{ad}$ is a local optimizer, and if $\tilde{J}$ is G-differentiable in u, then u satisfies

$$\langle D\tilde{J}(u), \tilde{u} - u\rangle_{U^*U} \geq 0 \quad \forall \tilde{u} \in U_{ad}. \tag{3.1}$$

If $\tilde{J}$ is convex and (3.1) holds for some $u \in U_{ad}$, then u is a global optimizer.

(iv) For the existence of a solution of Problem 3.2, the following conditions are sufficient: $\tilde{J}$ is convex and G-differentiable in a neighbourhood of U_{ad}, and (3.1) holds for some $u \in U_{ad}$.

Proof. See Subsections 1.2, 1.3 in Chapter 1 of [Lions, 1971], Theorem 1.46 in [Hinze et al., 2009]. □

By explicitly computing the derivative of $\tilde{J}$, these statements lead to the following conclusion on well-posedness of Problem 3.2.

Theorem 3.4. *Suppose that Assumption 3.1 holds, J and e are continuously F-differentiable and $e_y(y(u), u) \in \mathcal{B}(Y, Z)$ has a bounded inverse for each $u \in U$. Then $\tilde{J}$ is F-differentiable. Suppose further that*

1. *U is reflexive,*
2. *$\tilde{J}$ is strictly convex and $\tilde{J}(u) \to \infty$ when $\|u\|_U \to \infty$ in U_{ad},*
3. *U_{ad} is nonempty, closed and convex.*

Then there exists a unique solution $u \in U_{ad}$ to Problem 3.2 and it obeys, for some unique $p(u) \in Q^*$,

$$\begin{aligned} e(y(u), u) &= 0 && \text{in } Q, \\ e_y(y(u), u)^* p(u) + J_y(y(u), u) &= 0 && \text{in } Y^*, \\ \langle e_u(y(u), u)^* p(u) + J_u(y(u), u), \tilde{u}\rangle_{U^*U} &\geq 0 && \forall \tilde{u} \in U_{ad}. \end{aligned} \tag{3.2}$$

Proof. The proof is based on an application of the Implicit Function Theorem. See Section 1 in Chapter 2 of [Lions, 1971], or Sections 1.5.1 and 1.6.2 of [Hinze et al., 2009]. □

Remark. The conclusion of Theorem 3.4 also holds for more general forms of W_{ad} and in the case the maps J and e are only F-differentiable in a neighbourhood of U_{ad}. See [Hinze et al., 2009, Section 1.7] and [Carere, 2020, Section 1.3].

The first equation of (3.2) is again the state equation, while the second and third are known as the *adjoint equation* and *optimality criterion*, respectively. In the case of a linear-quadratic problem, we can now say the following.

Corollary 3.5. *Suppose U_{ad} is nonempty, closed and convex and that U is reflexive. In the case of a linear-quadratic problem, in which J and e are of the form (2.1) and (2.2), respectively, with A boundedly invertible, $M(z, z) \geq 0$ for $z \in Z$, and L coercive, there exists a unique minimizer $u \in U_{ad}$ of Problem 3.2 such that, for some unique $p(u) \in Q^*$,*

$$\begin{aligned} Ay(u) + Bu &= g && \text{in } Q, \\ C^*MCy(u) + A^*p(u) &= C^*Mz_d && \text{in } Y^*, \\ \langle Lu + B^*p(u), \tilde{u} - u\rangle_{U^*U} &\geq 0 && \forall \tilde{u} \in U_{ad}. \end{aligned} \tag{3.3}$$

Proof. For J being of quadratic form, F-differentiability is immediate. Its partial derivates can easily be computed, as M and L are self-adjoint, to be $\langle J_u(y,u),\tilde{u}\rangle_{U^*U} = \langle L\tilde{u},u\rangle_{U^*U}$ and $\langle J_y(y,u),\tilde{y}\rangle_{Y^*Y} = \langle MC\tilde{y},Cy\rangle_{Z^*Z} - \langle MC\tilde{y},z_d\rangle_{Z^*Z}$, for $\tilde{u}\in U,\tilde{y}\in Y$. Similarly, e is F-differentiable with $e_y = A, e_u = B$. We see that all partial derivatives are continuous, so that J and e are continuously F-differentiable. Clearly, $y(u) = A^{-1}(g-Bu)$ and Assumption 3.1 holds. The conditions on M and L imply $\tilde{J}(u)\geq\lambda\|u\|_U^2$, where λ denotes the coercivity constant of L. Hence $\tilde{J}(u)\to\infty$ as $u\to\infty$. By the same assumption on L it follows that $u\mapsto\langle Lu,u\rangle_{U^*U}$ is strictly convex. Furthermore, $u\mapsto\langle M(Cy(u)-z_d),Cy(u)-z_d\rangle_{Z^*Z}$ is convex by the assumption on M and the affinity of e. Strict convexity of $\tilde{J}$ follows. The result now follows from (3.2) by substitution. □

Remark. The assumptions for well-posedness usually considered in the literature often include that $Y = Q^*$ and A is coercive, i.e. $\sup_{0\neq y\in Y}\langle Ay,y\rangle_{QQ^*} > 0$, but well-posedness still holds if the first is omitted and if A is merely boundedly invertible. Note also that reflexivity of Y and Q is not required.

So to solve a linear-quadratic OCP under the conditions of Corollary 3.5, we must find the triple $(y,u,p)\in Y\times U\times Q^*$ that solves the system (3.3). In the *full admissibility* case, that is, in the case $W_{ad} = Y\times U$, the third equation becomes $Lu + B^*p = 0$ in U^*. Assuming also that Q is reflexive, so that $Q = P^*$ with $P := Q^*$, system (3.3) then can be written as

$$\begin{array}{llll} C^*MCy(u) & +A^*p(u) & = C^*Mz_d & \text{in } Y^*,\\ Lu & +B^*p(u) & = 0 & \text{in } U^*,\\ Ay(u) & +Bu & = g & \text{in } P^*. \end{array} \tag{3.4}$$

The Banach space P is called the *adjoint space*, and $p(u)$ the *adjoint solution*.

Remark. The obtained conclusions for linear-quadratic problems with full admissibility can also be recovered by studying saddle points of a Lagrangian, see e.g. [Kärcher and Grepl, 2014]. For a comparison between the Lagrangian approach and the adjoint approach presented here, see [Carere, 2020].

4 Truth Approximations

We present a standard approximation procedure for linear-quadratic OCPs with full admissibility based on Galerkin Projection. We assume that U and Q are reflexive and write $P = Q^*$. We also assume that $L(\boldsymbol{\mu})$ is coercive almost everywhere and that $\langle M(\boldsymbol{\mu})z,z\rangle_{Z^*Z}\geq 0$ for $z\in Z$ almost everywhere.

Taking closed subspaces $Y^{\mathcal{N}}\subset Y$, $U^{\mathcal{N}}\subset U$ and $P^{\mathcal{N}}\subset P$, we obtain for an outcome μ of $\boldsymbol{\mu}$, an approximation $(y^{\mathcal{N}}(\mu),u^{\mathcal{N}}(\mu),p^{\mathcal{N}}(\mu))\in Y^{\mathcal{N}}\times U^{\mathcal{N}}\times P^{\mathcal{N}}$ of $(y(\mu),u(\mu),p(\mu))$ by a Galerkin Projection of $(y(\mu),u(\mu),p(\mu))$ onto $Y^{\mathcal{N}}\times U^{\mathcal{N}}\times P^{\mathcal{N}}$. That is, we solve (3.4) with Y,U and P replaced by respectively $Y^{\mathcal{N}},U^{\mathcal{N}}$ and $P^{\mathcal{N}}$. Defining

$$\begin{aligned} A^{\mathcal{N}}(\boldsymbol{\mu})\in L^2(\Theta;\mathcal{B}(Y^{\mathcal{N}},(P^{\mathcal{N}})^*)), &\quad \text{by} \quad & A^{\mathcal{N}}(\boldsymbol{\mu})y^{\mathcal{N}} = (A(\boldsymbol{\mu})y^{\mathcal{N}})_{|P^{\mathcal{N}}},\\ B^{\mathcal{N}}(\boldsymbol{\mu})\in L^2(\Theta;\mathcal{B}(U^{\mathcal{N}},(P^{\mathcal{N}})^*)), &\quad \text{by} \quad & B^{\mathcal{N}}(\boldsymbol{\mu})u^{\mathcal{N}} = (B(\boldsymbol{\mu})u^{\mathcal{N}})_{|P^{\mathcal{N}}}, \end{aligned}$$

we notice that this amounts to solving, $\mathbb{P}^{\boldsymbol{\mu}}$-almost everywhere, the OCP

$$\begin{aligned} &\text{minimize } J_{|Y^{\mathcal{N}}\times U^{\mathcal{N}}}(\tilde{y}^{\mathcal{N}},\tilde{u}^{\mathcal{N}};\mu) \text{ s.t. } A^{\mathcal{N}}(\mu)\tilde{y}^{\mathcal{N}} + B^{\mathcal{N}}(\mu)\tilde{u}^{\mathcal{N}} = (g(\mu))_{|P^{\mathcal{N}}}\\ &\qquad\text{over all } (\tilde{y}^{\mathcal{N}},\tilde{u}^{\mathcal{N}})\in Y^{\mathcal{N}}\times U^{\mathcal{N}}. \end{aligned}$$

This approximate problem is called the *truth problem*. By Corollary 3.5, it is well-posed if $A^{\mathcal{N}}(\mu)$ is boundedly invertible, as all other conditions are inherited from the original OCP in the continuous formulation. In that case, the approximate solution $(y^{\mathcal{N}}(\mu),u^{\mathcal{N}}(\mu),p^{\mathcal{N}}(\mu))$ is called the *truth approximation*. In Section 6 we shall take $Y^{\mathcal{N}}$, $U^{\mathcal{N}}$ and $P^{\mathcal{N}}$ to be Finite Element spaces of dimension of order $O(\mathcal{N})$, see e.g. [Quarteroni and Valli, 1994, Chapters 3,5,6].

Remark. Just as for well-posedness of the original OCP it is not necessary for Y and P to be equal, it is not required for $Y^{\mathcal{N}}$ and $P^{\mathcal{N}}$ to be equal, as long as $A^{\mathcal{N}}(\mu)$ is boundedly invertible.

5 Reduced Order Method

In addition to the assumptions of Section 4, let us now assume that Y, U and Q are Hilbert spaces, so that also P is a Hilbert space. If $\mu \in \mathbb{M}$ is an outcome of $\boldsymbol{\mu}$, then we could compute the expensive truth approximation $(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu))$ by performing a Galerkin Projection of $(y(\mu), u(\mu), p(\mu))$ onto the subspace $Y^{\mathcal{N}} \times U^{\mathcal{N}} \times P^{\mathcal{N}}$, which has a high dimension of order $O(\mathcal{N})$. This high-dimensionality is necessary to ensure the accuracy of the truth approximations. Since solving many truth problems becomes computationally infeasible, in this section we propose the construction of a cheap *reduced basis approximation* $(y_N(\mu), u_N(\mu), p_N(\mu))$ using a ROM, [Haasdonk, 2017, Hesthaven et al., 2016]. We discuss the implementation of ROMs for OCPs as already presented in [Bader et al., 2017, Dedè, 2010, Kärcher et al., 2018, Negri et al., 2015, Negri et al., 2013].

5.1 Reduced basis approximation

Given $\mu \in \mathbb{M}$, the reduced basis approximation $(y_N(\mu), u_N(\mu), p_N(\mu))$ can be obtained by performing a Galerkin Projection of $(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu))$ onto a subspace $\mathcal{X}_N$ of $Y^{\mathcal{N}} \times U^{\mathcal{N}} \times P^{\mathcal{N}}$, which has a low dimension of order $O(N)$, with $N \ll \mathcal{N}$. In this subsection we discuss the construction of such a space based on the approach known as *weighted Proper Orthogonal Decomposition* (*weighted POD*), or just *Proper Orthogonal Decomposition* (*POD*). The process of this construction is called the *offline phase*, and the resulting low dimensional space is called the *reduced basis space*. The act of performing the Galerkin Projection onto the reduced basis space for a given outcome μ of $\boldsymbol{\mu}$ is known as the *online phase*.

Given such an outcome $\mu \in \mathbb{M}$, the solution to the OCP $(y(\mu), u(\mu), p(\mu))$ can be approximated by the truth solution $(y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu))$, which, in turn, can be approximated with the reduced basis approximation denoted by $(y_N(\mu), u_N(\mu), p_N(\mu))$. In order for the ROM to have any use, one should be able to construct arbitrarily precise truth approximations, uniformly in the parameter. The whole task then is to construct a ROM in such a way that it is effective, i.e. that the error in approximating the truth solution by the reduced basis approximation is also small, and that $(y_N(\mu), u_N(\mu), p_N(\mu)$ can be computed efficiently.

5.2 Offline phase: weighted Proper Orthogonal Decomposition

Let us write $\mathcal{X} = Y \times U \times P$. Assuming $\chi^{\mathcal{N}}(\boldsymbol{\mu}) := (y^{\mathcal{N}}(\boldsymbol{\mu}), u^{\mathcal{N}}(\boldsymbol{\mu}), p^{\mathcal{N}}(\boldsymbol{\mu}) \in L^2(\Theta; \mathcal{X})$, in order to find a suitable low dimensional space of $\mathcal{X}^{\mathcal{N}} := Y^{\mathcal{N}} \times U^{\mathcal{N}} \times P^{\mathcal{N}}$, one could choose to construct the subspace $\mathcal{X}_N \subset \mathcal{X}^{\mathcal{N}}$ in such a way that it minimizes the expected squared error

$$\mathbb{E}\|\chi^{\mathcal{N}}(\mu) - P_V \chi^{\mathcal{N}}(\mu)\|^2_{\mathcal{X}} = \int_{\mathbb{M}} \|\chi^{\mathcal{N}}(\mu) - P_V \chi^{\mathcal{N}}(\mu)\|^2_{\mathcal{X}} \, \mathrm{d}\mathbb{P}^{\boldsymbol{\mu}}(\mu),$$

among all subspaces V of $\mathcal{X}^{\mathcal{N}}$ of dimension at most N. Here P_V denotes the orthogonal projector onto V.

Recall that the (discrete) solution manifold $\mathcal{M}^{\mathcal{N}}$ is defined as $\{\chi^{\mathcal{N}}(\mu), \mu \in \mathbb{M}\}$. Defining $\mathcal{C} \in \mathcal{B}(\mathrm{span}\{\mathcal{M}^{\mathcal{N}}\}, \mathrm{span}\{\mathcal{M}^{\mathcal{N}}\})$ to be the compact, self-adjoint and positive operator

$$\mathcal{C}v = \mathbb{E}\left[\langle v, \chi^{\mathcal{N}}(\boldsymbol{\mu})\rangle_{\mathcal{X}} \, \chi^{\mathcal{N}}(\boldsymbol{\mu})\right], \tag{5.1}$$

it is well known (e.g. [Schwab and Todor, 2006, Griebel and Harbrecht, 2018]) that $\mathcal{X}_N = \mathrm{span}\{\xi_1, \ldots, \xi_N\}$, where $(\lambda_i, \xi_i)_i$ is the eigenvalue-eigenvector sequence of $\mathcal{C}$ in which the eigenvalues are ordered in a decreasing fashion. Notice that the eigenvalues are positive and can accumulate only in 0. The problem is that this expectation is in general not known. To circumvent this problem, the weighted POD method uses a quadrature rule to approximate it.

Let $\mathbb{M}_d := \{\mu_1, \ldots, \mu_M\} \subset \mathbb{M}$ and $\{w_1, \ldots, w_M\} \subset \mathbb{R}$ denote, respectively, the nodes and weights of a quadrature rule for $\mathbb{P}^{\boldsymbol{\mu}}$. The subscript '$d$' stands for 'discrete'. We admit only nonzero weights, and write χ_i for the ith *snapshot* $\chi^{\mathcal{N}}(\mu_i)$. Let us define $\mathcal{C}_d \in \mathcal{B}(\mathrm{span}\{\chi_1, \ldots, \chi_M\}, \mathrm{span}\{\chi_1, \ldots, \chi_M\})$, the discrete counterpart of $\mathcal{C}$, as

$$\mathcal{C}_d v = \sum_{i=1}^{M} w_i \langle v, \chi_i \rangle_{\mathcal{X}} \, \chi_i.$$

This operator is compact and self-adjoint but positive definite only when all weights are positive. It is not difficult to show, however, that the space $\mathcal{X}_N$ that minimizes the approximate

$$\sum_{i=1}^{M} w_i \|\chi_i - P_V \chi_i\|_{\mathcal{X}}^2,$$

of (5.1) over all subspaces V of $\mathcal{X}^{\mathcal{N}}$ of dimension at most N, is given by $\mathcal{X}_N = \operatorname{span}\{\xi_1, \dots, \xi_{N \wedge K}\}$, where $(\lambda_i, \xi_i)_i$ now is the eigenpair sequence of $\mathcal{C}_d$ in which the eigenvalues are ordered in a decreasing fashion and K is the number of positive eigenvalues, see e.g. [Venturi, 2016, Section 1.2].

In a numerical implementation of the weighted POD, the operator $\mathcal{C}_d$ is expressed in a basis. Two possibilities for this are

- expressing the operator $\mathcal{C}_d$ in the snapshot basis $(\chi_i)_{i=1}^M$. Let us denote the resulting matrix by C, so $C_{ij} = w_i \langle \chi_j, \chi_i \rangle_{\mathcal{X}}$. Define also the positive definite, symmetric matrix $C_0 = (\frac{1}{M} \langle \chi_j, \chi_i \rangle_{\mathcal{X}})_{ij}$ and the weight matrix $P = \operatorname{diag}(w_1, \dots, w_M)$. To get the basis $(\xi_i)_{i=1}^N$ of the minimizing subspace $\mathcal{X}_N$, the N leading orthonormalized eigenvectors $(\boldsymbol{\xi}_i)_{i=1}^N$ of $C = MPC_0$ are computed[1]. As these are expressed in the snapshot basis, we expand each of them in this snapshot basis to obtain $(\xi_i)_{i=1}^N$. That is, the ith basis function is $\xi_i = \sum_{j=1}^M (\boldsymbol{\xi}_i)_j \chi_j$.
- Another possibility, when the weights are all positive, is to express $\mathcal{C}_d$ in the weighted snapshot basis $(\sqrt{w_i}\chi_i)_{i=1}^M$. We denote the resulting matrix as C_w, so $(C_w)_{ij} = \sqrt{w_i}\sqrt{w_j}\langle \chi_j, \chi_i \rangle_{\mathcal{X}}$. Notice that C_w is positive definite and symmetric. The N leading orthonormalized eigenvectors $(\boldsymbol{\xi}_i^w)_{i=1}^N$ of C_w are computed. As they are expressed in the weighted snapshot basis, the functions $(\xi_i)_{i=1}^N$ can be recovered by expanding $(\boldsymbol{\xi}_i^w)_{i=1}^N$ in this weighted snapshot basis: $\xi_i = \sum_{j=1}^M (\boldsymbol{\xi}_i^w)_j \sqrt{w_j} \chi_j$.

The weighted POD thus picks out the most important directions, according to an L^2-type criterion, of $\operatorname{span}\{\chi_1, \dots, \chi_M\}$. As only one solution manifold is involved, this is known as the *monolithic approach*. Instead, we shall perform separate weighted PODs on each of the solution manifolds

$$\{y^{\mathcal{N}}(\mu_1), \dots, y^{\mathcal{N}}(\mu_M)\}, \quad \{u^{\mathcal{N}}(\mu_1), \dots, u^{\mathcal{N}}(\mu_M)\}, \quad \{p^{\mathcal{N}}(\mu_1), \dots, p^{\mathcal{N}}(\mu_M)\}.$$

The resulting low dimensional spaces Y_N, U_N and P_N are combined to furnish the reduced basis space $\mathcal{X}_N = Y_N \times U_N \times P_N$. This approach, known as the *partitioned approach*, has been observed to be preferable to the monolithic approach in a variety of scenarios, by [Strazzullo et al., 2017]. Of course, if a solution manifold is already contained in a subspace of dimension at most N, say the solution manifold of the control $\{u^{\mathcal{N}}(\mu_1), \dots, u^{\mathcal{N}}(\mu_M)\}$, no POD is needed and one simply takes the linear span of this manifold as U_N. If Y, U, or P is a product of Hilbert spaces itself, more POD compressions could also be performed, for example one for each space in this product.

Remark. The term "POD" is often reserved for a weighted POD in which a Monte Carlo sampler is used as quadrature rule, see [Venturi et al., 2019a, Venturi et al., 2019b].

Remark. Some quadrature rules, such as Gaussian quadrature, rely on appropriate smoothness of the map $\mu \mapsto \chi^{\mathcal{N}}(\mu)$.

5.3 Online phase

The online phase now consists of solving the system (3.4) with Y, U, P replaced by Y_N, U_N, P_N, respectively, for any $\mu \in \mathbb{M}$ of interest. That is, we perform a Galerkin projection of $(y(\mu), u(\mu), p(\mu))$ onto $Y_N \times U_N \times P_N$. This amounts to solving a system of size $3N \times 3N$, if $\dim Y_N = \dim U_N = \dim P_N = N$. It should be noted that for this reduced problem to be well-posed for ($\mathbb{P}^{\boldsymbol{\mu}}$-almost) every $\mu \in \mathbb{M}$, the operator $A_N(\mu) \in \mathcal{B}(Y_N, (P_N)^*)$ defined by $A_N(\mu)\tilde{y}_N = (A(\mu)\tilde{y})_{|Y_N}$ for $\tilde{y}_N \in Y_N$, should be boundedly invertible for ($\mathbb{P}^{\boldsymbol{\mu}}$-almost) every $\mu \in \mathbb{M}$.

Remark. It is customary to define an *aggregate space* or *integrated space* $Z_N = \operatorname{span}\{Y_N, P_N\}$ and put $\mathcal{X}_N = Z_N \times U_N \times Z_N$ as reduced basis space instead. In that case the subspaces of Y and P coincide but the system to solve online is of the larger size $5N \times 5N$. We stress that it is not fundamental to perform this step in order for the OCP to be well-posed. It *is* fundamental that $A_N(\mu)$ is boundedly invertible for

[1] Instead of solving $MPC_0 x = \lambda x$, it is preferable to solve $MC_0 x = \lambda P^{-1} x$, as P^{-1} can be easily computed and more efficient solvers are available when P is positive definite.

every μ, as follows from the same argumentation of Section 4. If $A(\mu)$ happens to be coercive, then one can use aggregate spaces to inherit the coercivity and hence invertibility of $A_N(\mu)$ from $A(\mu)$. However, in a noncoercive setting, the invertibility of $A_N(\mu)$ should be guaranteed in a different way, e.g. by using supremizer solutions, see [Rozza et al., 2013, Rozza and Veroy, 2007].

In a numerical implementation of the online phase, the optimality system must be expressed in a basis of $\mathcal{X}_N$. If the computational complexity of this operation is independent of $\mathcal{N}$, then so is the complexity of the online phase. The problem is that such a basis is itself expressed in the $\mathcal{N}$-dimensional snapshot basis, so that this can not be guaranteed without additional assumptions. To this end, we require that $A(\boldsymbol{\mu}), B(\boldsymbol{\mu}), M(\boldsymbol{\mu}), L(\boldsymbol{\mu}), g(\boldsymbol{\mu})$ and $z_d(\boldsymbol{\mu})$ all adhere to the following separation of the variables.

Assumption 5.1. We assume the following *affine decompositions* hold (for $\mathbb{P}$-almost every μ):

$$
\begin{aligned}
A(\mu) &= \sum_{q=1}^{Q_A} \Lambda_A^q(\mu) A_q, & B(\mu) &= \sum_{q=1}^{Q_B} \Lambda_B^q(\mu) B_q, \\
M(\mu) &= \sum_{q=1}^{Q_M} \Lambda_M^q(\mu) M_q, & L(\mu) &= \sum_{q=1}^{Q_L} \Lambda_L^q(\mu) L_q, \\
g(\mu) &= \sum_{q=1}^{Q_g} \Lambda_g^q(\mu) g_q, & z_d(\mu) &= \sum_{q=1}^{Q_{z_d}} \Lambda_{z_d}^q(\mu) (z_d)_q,
\end{aligned}
$$

where

- $A_q \in \mathcal{B}(Y, P^*),\ B_q \in \mathcal{B}(U, P^*),\ M_q \in \mathcal{B}(Z, Z^*),\ L_q \in \mathcal{B}(U, U^*),\ g_q \in P^*,\ z_d^q \in Z,$
- $\Lambda_A^q, \Lambda_B^q, \Lambda_M^q, \Lambda_L^q, \Lambda_g^q, \Lambda_{z_d}^q : \mathbb{M} \to \mathbb{R},$
- $Q_A, Q_B, Q_M, Q_L, Q_g, Q_{z_d} \in \mathbb{N}.$

Further argumentation can be found in [Hesthaven et al., 2016, Section 3.3]. If the above parametric maps are continuous, then under some additional assumptions on the involved operators the solution maps

$$
\begin{aligned}
&\mu \mapsto (y(\mu), u(\mu), p(\mu)), \\
&\mu \mapsto (y^{\mathcal{N}}(\mu), u^{\mathcal{N}}(\mu), p^{\mathcal{N}}(\mu)), \\
&\mu \mapsto (y_N(\mu), u_N(\mu), p_N(\mu)),
\end{aligned}
$$

are continuous as well, and are therefore measurable, bounded, and in $L^2(\mathbb{M}; Y \times U \times P)$, see Proposition 2.2.6 in [Carere, 2020].

6 Numerical Applications

In this section we extend the three environmental applications, initially introduced in [Strazzullo, 2016, Strazzullo et al., 2017], by modelling the random variable $\boldsymbol{\mu}$ to have a distribution other than the uniform distribution, and assess the results of the ROMs constructed by a weighted POD approach. In these examples, the choice of training and testing set sizes are based on [Strazzullo et al., 2017]. In the first application the spill of hypothetical pollutant in a fluid described by an elliptic PDE is studied. The second and third examples consist of an ocean circulation model, which is based on the quasi-geostrophic equations which form a noncoercive PDE. In these latter two numerical examples, a comparison between the results obtained with and without the usual aggregation procedure is given, but we shall not be too concerned with the well-posedness of the OCPs in these two examples.

6.1 Hypothetical pollution in the Gulf of Trieste

We model a hypothetical spill of a pollutant near the harbor of Koper, Slovenia, in the Gulf of Trieste, Italy. It is governed by an elliptic steady advection-diffusion equation.

Problem Formulation The domain $\Omega \subset \mathbb{R}^2$ with Lipschitz boundary is an approximation of the geographical area of the Gulf of Trieste. A fine triangulation $\mathcal{T}_h$ of mesh size $h > 0$ has been constructed and shared by the authors of [Strazzullo et al., 2017]. In Figure 6.1 (*left/center*) this domain is shown, as well as $\mathcal{T}_h$. Two subdomains of Ω are of particular importance, the subdomain of the spill Ω_u, and the domain of observation, Ω_{obs}.

Ω_u the area in which the pollutant is spilled, corresponding to the geographical area of the harbor of Koper, Slovenia,

Ω_{obs} the domain of observation, the Miramare natural reserve, which is of interest being a protected environment due to its ecological flora and fauna and being a prominent area to relax for tourists and the citizens of Trieste.

The boundary $\Gamma := \partial\Omega$ is subdivided in a Dirichlet boundary Γ_D and a Neumann boundary Γ_N, where $\Gamma = \Gamma_D \cup \Gamma_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$, and on which homogeneous Dirichlet and Neumann boundary conditions are imposed, respectively. The Dirichlet boundary corresponds to the coastal part of Γ while the Neumann boundary corresponds to open sea, see Figure 6.1.

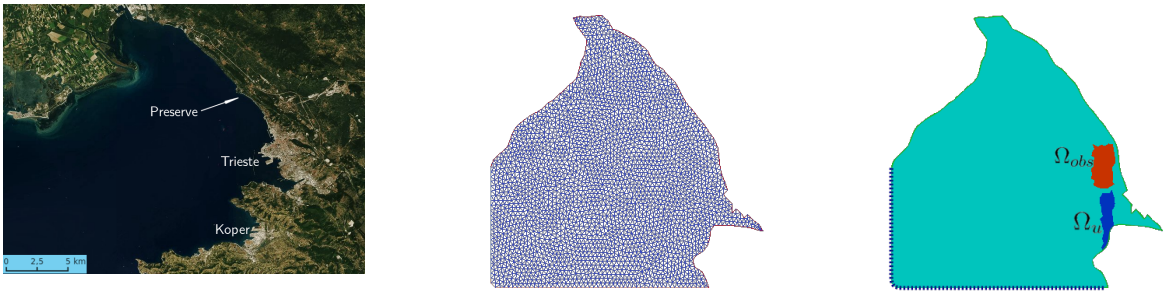

Figure 6.1: *left*: geographical area of the Gulf of Trieste, *center*: physical domain Ω with triangulation, *right*: physical domain with subdomains Ω_u and Ω_{obs}, and sub boundaries Γ_N (dotted blue) and Γ_D (solid green).

Let us denote the state variable y as the pollutant concentration, while the desired concentration $z_d = 0.2\chi_{\Omega_{obs}} \in L^2(\Omega)$ represents the maximal safe concentration of pollutant on Ω_{obs}, where $\chi_{\Omega_{obs}}$ is the indicator of Ω_{obs}. The natural state space is $Y := H^1_{\Gamma_D}(\Omega) = \left\{v \in H^1(\Omega) : \ v = 0 \text{ on } \Gamma_D\right\}$. For $u \in U := \mathbb{R}$, consider $u_0 = u\chi_{\Omega_u} \in L^2(\Omega)$. The function u_0 models a source on Ω_u originating from the spill. We are interested in finding the value of u for which the pollutant concentration in Ω_{obs} equals the maximal safe concentration z_d, or is as close to z_d as possible in an $L^2(\Omega)$ sense.

The problem above can be phrased as the following OCP, where the governing advection-diffusion state equation is written directly in its weak form:

$$
\begin{aligned}
&\text{minimize } J(y,u) = \frac{1}{2}\int_{\Omega_{obs}} |y - z_d|^2 \,\mathrm{d}x + \frac{\alpha}{2}\int_{\Omega_u} |u|^2 \,\mathrm{d}x \text{ over all } (y,u) \in Y \times U,\\
&\text{such that } A(\mu)y + B(\mu)u = 0, \text{ where}\\
&\langle A(\mu)y, \tilde{p}\rangle_{P^*P} = \int_\Omega \mu_1 \nabla y \cdot \nabla \tilde{p} \,\mathrm{d}x + \int_\Omega ([\mu_2, \mu_3] \cdot \nabla y)\, \tilde{p} \,\mathrm{d}x,\\
&\langle B(\mu)u, \tilde{p}\rangle_{P^*P} = -L_0 u \int_{\Omega_u} \tilde{p} \,\mathrm{d}x,\\
&\text{and } P = Y.
\end{aligned}
$$

We take $\mathbb{M} = [\frac{1}{2}, 1] \times [-1, 1] \times [-1, 1]$. The parameter μ describes specific properties of the sea: μ_1 is a diffusivity parameter while $[\mu_2, \mu_3]$ models a constant advective field. The constant $L_0 = 1000$ is used to put the state equation in nondimensional form. We are mostly interested in minimizing the first term in the objective functional. On the other hand, to ensure uniqueness of the optimal control, we do prescribe a small positive value for α, namely $\alpha = 10^{-7}$. When $A(\mu)$ is coercive for every μ, it is in

particular invertible for every $\mu \in \mathbb{M}$ so that the OCP is well-posed for every μ by Corollary 3.5, and to find the solution we must solve (3.4). Notice that in the formulation of Definition 2.2, the observation space Z is taken to be $L^2(\Omega)$, M and L/α are the Riesz map on $L^2(\Omega)$, C is the injection $Y \hookrightarrow Z$ and $g = 0$.

Coercivity of $A_N(\mu)$ Coercivity of $A_N(\mu)$ holds if the Poincaré and Trace constants C_p and C_t, given in the inequalities

$$C_p = \inf\left\{c > 0 : \int_\Omega |u|^2 \,\mathrm{d}x \leq c \int_\Omega |\nabla u|^2 \,\mathrm{d}x \quad \forall u \in H^1_{\Gamma_D}(\Omega)\right\}, \tag{6.1}$$
$$C_t = \inf\left\{c > 0 : \int_{\Gamma_N} |u|^2 \,\mathrm{d}x \leq c \int_\Omega |u|^2 + |\nabla u|^2 \,\mathrm{d}x \quad \forall u \in H^1_{\Gamma_D}(\Omega)\right\},$$

are small enough: $(C_p + 1)C_t < \frac{1}{2}\sqrt{2}$. In order to see this, notice that Γ_N consists of a western vertical part W, a southern horizontal part S, and a small diagonal part SW angled at 45 degrees that joins W and S. With a convective field $[\mu_2, \mu_3]$ and outer normal n at the boundary, we have on Γ_N that

$$[\mu_2, \mu_3] \cdot n = \mu_2 \chi_W + \mu_3 \chi_S + \frac{1}{2}\sqrt{2}(\mu_2 + \mu_3)\chi_{SW} \geq -\chi_W - \chi_E - \sqrt{2}\chi_{SW} \geq -\sqrt{2}\chi_{\Gamma_N},$$

Writing $y\nabla y = \frac{1}{2}\nabla y^2$ and using Green's Theorem and the Trace and Poincaré inequalities, we obtain for arbitrary $(\mu_1, \mu_2, \mu_3) \in \mathbb{M}$ and $y \in H^1_{\Gamma_D}(\Omega)$

$$\begin{aligned}
\langle A(\mu)y, y\rangle_{Y^*Y} &= \mu_1 \int_\Omega |\nabla y|^2 \,\mathrm{d}x + \frac{1}{2}\int_{\Gamma_N} ([\mu_2, \mu_3] \cdot n) y^2 \,\mathrm{d}x, \\
&\geq \frac{1}{2}\int_\Omega |\nabla y|^2 \,\mathrm{d}x - \frac{1}{2}\sqrt{2}\int_{\Gamma_N} |y|^2 \,\mathrm{d}x \\
&\geq \frac{1}{2}\int_\Omega |\nabla y|^2 \,\mathrm{d}x - \frac{1}{2}\sqrt{2}C_t \int_\Omega (|y|^2 + |\nabla y|^2) \,\mathrm{d}x \\
&\geq \frac{1}{2}\int_\Omega |\nabla y|^2 \,\mathrm{d}x \left(1 - \sqrt{2}C_t - \sqrt{2}C_t C_p\right).
\end{aligned}$$

Since by again the Poincaré inequality (6.1) the $H^1(\Omega)$-seminorm $y \mapsto \int_\Omega |\nabla y|^2 \,\mathrm{d}x$ is equivalent to the $H^1(\Omega)$-norm on $H^1_{\Gamma_D}(\Omega)$, we see that coercivity is ensured if $(C_p + 1)C_t < \frac{1}{2}\sqrt{2}$.

Truth Approximation The high fidelity spaces are based on the Finite Elements

$$X^k_h = \left\{v \in C^0(\Omega) : \ v_{|K} \in P_k(K) \ \forall K \in \mathcal{T}_h\right\} \subset H^1(\Omega), \tag{6.2}$$

where $P_k(K)$ is the space of polynomials on the element K of degree at most k. More precisely, as the high fidelity state, control, and adjoint spaces $Y^{\mathcal{N}}, U^{\mathcal{N}}, P^{\mathcal{N}}$ we take

$$Y^{\mathcal{N}_Y} = X^1_h \cap Y, \quad U^{\mathcal{N}_U} = \mathbb{R}, \quad P^{\mathcal{N}_P} = X^1_h \cap Y.$$

The trace and Poincaré inequalities remain valid on $Y^{\mathcal{N}}$, with the (smaller) corresponding constants $C_t^{\mathcal{N}}$ and $C_p^{\mathcal{N}}$. By solving an eigenvalue problem we computed these constants to be $C_t^{\mathcal{N}} = 0.52$ and $C_p^{\mathcal{N}} = 0.06$. Hence, coercivity of $A^{\mathcal{N}}(\mu)$ is guaranteed for every $\mu \in \mathbb{M}$ since we take $Y^{\mathcal{N}} = P^{\mathcal{N}}$. The truth problem described in Section 4 thus is well-posed by Corollary 3.5. Furthermore, the truth solution can be shown to converge to the solution of the problem in continuous formulation uniformly on $\mathbb{M}$ (see [Carere, 2020]).

Reduced Order Model In this case we need only perform a POD compression on state and adjoint, and can leave $U_N = U$. We do aggregate state and adjoint in this application, so that $A_N(\mu)$ inherits the coercivity of $A(\mu)$ for every μ. If $\dim Y_N = \dim P_N = N$, then it remains to solve a system of size $(4N + 1) \times (4N + 1)$, which now is well-posed due to the inherited coercivity for $A_N(\mu)$ for each μ.

Because L, M, B, g and z_d are parameter independent and A can be decomposed in the $Q_A = 3$ terms

$$\Lambda_A^1(\mu) = \mu_1, \qquad \langle A_1 y, \tilde{y}\rangle_{Y^*Y} = \int_\Omega \nabla y \cdot \nabla \tilde{y}\,\mathrm{d}x,$$
$$\Lambda_A^2(\mu) = \mu_2, \qquad \langle A_2 y, \tilde{y}\rangle_{Y^*Y} = \int_\Omega \frac{\partial y}{\partial x_1}\tilde{y}\,\mathrm{d}x,$$
$$\Lambda_A^3(\mu) = \mu_3, \qquad \langle A_3 y, \tilde{y}\rangle_{Y^*Y} = \int_\Omega \frac{\partial y}{\partial x_2}\tilde{y}\,\mathrm{d}x.$$

Assumption 5.1 holds, hence this system can be solved in a computation count independent of $\mathcal{N}$.

Training set generation As mentioned in Section 5, the weighted POD is based on the choice of a quadrature rule with nodes $\mathbb{M}_d = \{\mu_1, \dots \mu_M\}$ and weights $\{w_1, \dots, w_M\}$. Writing $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3)$, we shall only consider product measures $\mathbb{P}^{\boldsymbol{\mu}} = \mathbb{P}^{\boldsymbol{\mu}_1} \times \mathbb{P}^{\boldsymbol{\mu}_2} \times \mathbb{P}^{\boldsymbol{\mu}_3}$ for which each $\mathbb{P}^{\boldsymbol{\mu}_i}$ admits a Lebesgue density, and we will assess the performance of the following quadrature rules (see e.g. [Sullivan, 2015]):

- a Monte Carlo sampler, which samples the nodes from $\mathbb{P}^{\boldsymbol{\mu}}$. Its weights are all equal to $\frac{1}{M}$.
- The tensor product of three Gaussian quadrature rules. In this case the weights are not all equal.
- A Pseudo-Random sampler, that provides a rule for $\mathrm{Un}([0,1] \times [0,1] \times [0,1])$, the Uniform distribution on $[0,1] \times [0,1] \times [0,1]$. To get a rule for $\mathbb{P}^{\boldsymbol{\mu}}$ on $\mathbb{M}$, we use the method of inversion for each $\mathbb{P}^{\boldsymbol{\mu}_i}$.
- A tensor product of Clenshaw-Curtis quadrature rules, which provide[2] a rule for $\mathrm{Un}([-1,1])$. To get a rule for $\mathbb{P}^{\boldsymbol{\mu}}$, we use a change of variables for each i.

Results An image of the truth solution for the state and adjoint component is shown in Figure 6.2. The parameter value for which this solution is obtained is $\mu = (1, -1, 1)$. This solution is thus obtained with a convective field $(\mu_2, \mu_3) = (-1, 1)$ which models a water flow from south-east to north-west. The value of the optimal control is 7.4×10^{-1}. Notice that the truth solution for the state is strictly

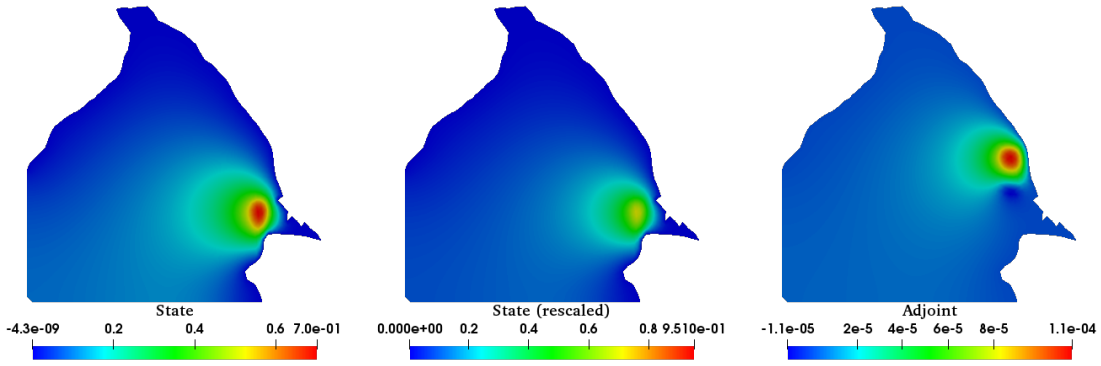

Figure 6.2: Gulf: truth solution for state (*left*, *center*) and adjoint (*right*), for $\mu = (1, -1, 1)$.

smaller than z_d on Ω_{obs}. To enable visual comparison with the result of [Strazzullo et al., 2017], the state solution is also displayed with a rescaled color range. The dimension of the truth problem is[3] 5345×5345.

Supposing first that $\boldsymbol{\mu}$ is uniformly distributed, we build ROMs with $N = 1, \dots, 35$, which are constructed with a Monte Carlo sampling procedure to obtain a training set $\mathbb{M}_d$ of size 100. For $\mu = (1, -1, 1)$, a plot of the pointwise difference between the reduced state and adjoint solutions obtained for $N = 35$ and the respective truth solutions is shown in Figure 6.3.

[2] The Clenshaw-Curtis quadrature rule can also be implemented to provide a rule for different densities.

[3] To be precise, we are reporting the total number of nonzero coefficients of the expansion of the truth solution components $y^{\mathcal{N}}, p^{\mathcal{N}}$ for $\mu = (1, -1, 1)$ in the Finite Elements basis (6.2).

Next, we consider the relative error for state, control and adjoint solution components and the output functional, i.e.

$$e_{y,N}(\mu) = \frac{\|y^{\mathcal{N}}(\mu) - y_N(\mu)\|_Y}{\|y^{\mathcal{N}}(\mu)\|_Y},$$
$$e_{u,N}(\mu) = \frac{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_U}{\|u^{\mathcal{N}}(\mu)\|_U},$$
$$e_{p,N}(\mu) = \frac{\|p^{\mathcal{N}}(\mu) - p_N(\mu)\|_P}{\|p^{\mathcal{N}}(\mu)\|_P},$$
$$e_{J,N}(\mu) = \frac{|J^{\mathcal{N}}(\mu) - J_N(\mu)|}{|J^{\mathcal{N}}(\mu)|}.$$

We plot their base-10 logarithm, averaged over values of μ in a testing set of size 100 which is sampled from $\mathbb{P}^{\boldsymbol{\mu}}$ and is different from the training set, in Figure 6.4. The sample average gives an indication of the trend of the decay. Not all parameters follow this trend exactly. For parameter value for which the problem is inherently more difficult to solve, the error decays more slowly. The amount of variation of the logarithmic relative errors among the parameters in the training set is indicated in the plots, by including the sample standard deviation. For example, for the state this is an unbiased estimator of the true standard deviation

$$\sqrt{\int_\Theta \left(\log_{10}\|e_{y,N}(\boldsymbol{\mu})\|_Y - \mathbb{E}\log_{10}\|e_{y,N}(\boldsymbol{\mu})\|_Y\right)^2 \mathrm{d}\mathbb{P}}.$$

The same holds true for the control and adjoint components and the output functional.

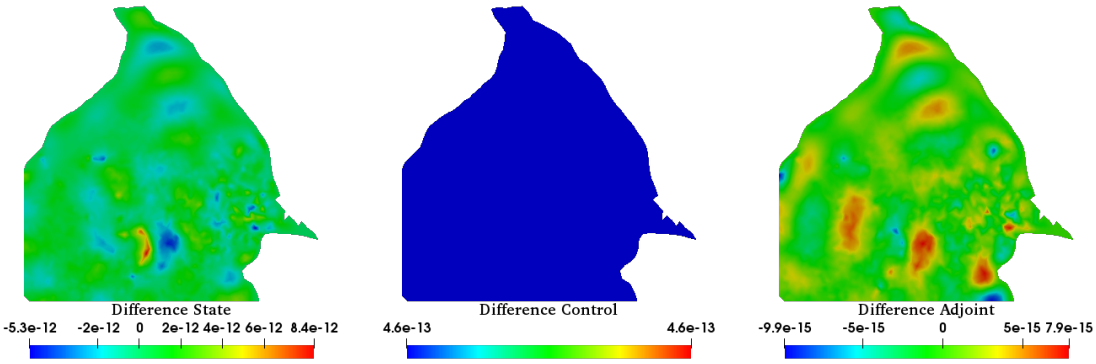

Figure 6.3: Gulf: ROM obtained with Monte Carlo sampling from uniform distribution: pointwise difference between reduced and truth state, control and adjoint solutions for $\mu = (1, -1, 1)$ (from left to right).

The pointwise errors for state, control and adjoint are very small and the relative normed errors decay exponentially. The ROM corresponding to $N = 35$ reaches a high accuracy of order 10^{-11} for the state variable[4].
The different quadrature rules specified above are implemented as well. In Figure 6.5 (*left*) the state error originating from ROMs that are constructed with Clenshaw-Curtis and Gaussian quadrature rules as specified above, are compared with the Monte Carlo sampler. A comparison is also made with a Pseudo Random sampler. The size of the training sets for the Monte Carlo and Pseudo Random samplers is 100. For each of the three parameters, a Gaussian and Clenshaw-Curtis quadrature of five nodes is taken, leading to tensor product rules with a training set size of 125. All four quadrature rules perform similarly. Finally, $\boldsymbol{\mu}$ is given a Beta(75,75) distribution for all three parameters, which puts most probability mass around the center of the parameter domain. The ROM is effectively using the additional information, as it requires only half the number of reduced functions compared to the deterministic case. This can be seen in Figure 6.5 (*right*). The ROMs, constructed with the Pseudo Random, Gauss and Monte Carlo rule, all perform similarly. Having thus concluded that the proposed ROM is able to make use of a low

[4]After $N = 35$, the error typically increases again due to numerical error, originating from the normalization procedure after the weighted POD, which is large for eigenvectors corresponding to the very small eigenvalues. We shall only plot the phase of decay.

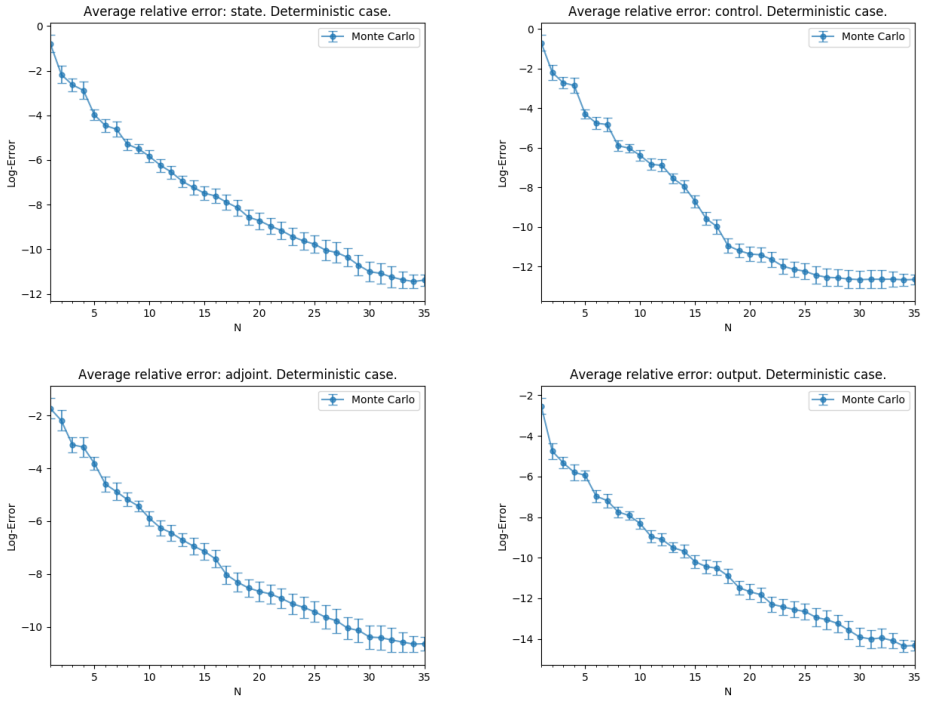

Figure 6.4: Gulf: average logarithmic relative errors for state (*top left*), control (*top right*), adjoint (*bottom left*) and output (*bottom right*) in the deterministic case, as a function of N. ROM obtained via Monte Carlo sampling.

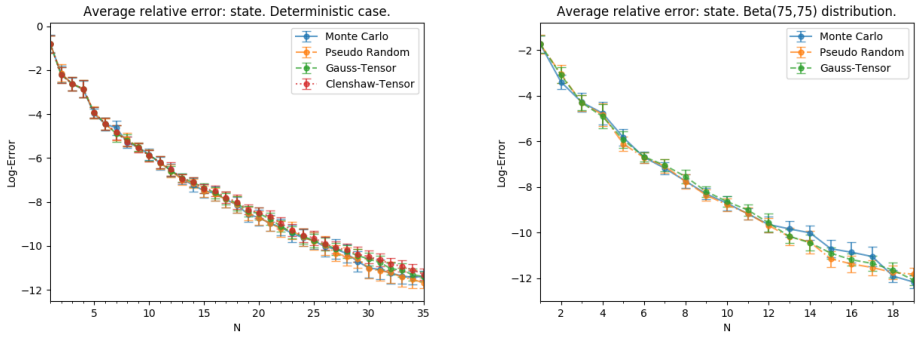

Figure 6.5: Gulf: average logarithmic relative errors for the state in the deterministic case (*left*), and under a Beta(75,75) distribution (*right*), as a function of N. Various quadrature rules are compared.

N	average	min	max	deviation
2	52.8	26.8	81.5	6.3
7	51.4	23.7	72.1	6.7
14	46.1	18.6	69.7	7.4
21	42.6	19.1	63.3	5.8
28	35.9	19.2	50.5	5.1
35	31.2	18.5	45.7	4.1

Table 6.1: Gulf: sample average, minimal value, maximal value, and sample standard deviation of speedup-index obtained with a testing set of size 100. Deterministic case. ROM obtained by Monte Carlo sampling.

dimensional solution space which remains very accurate, we further comment on efficiency. To verify it, for each μ in the testing set we compute the so-called *speedup-index*:

$$\text{speedup-index} = \frac{\text{computation time of truth solution}}{\text{computation time of reduced basis approximation}}.$$

Table 6.1 lists[5], for a few values of N, the sample average and sample standard deviation of the computed speedup-indices over the testing set of size 100, in the deterministic case. The minimal and maximal speedup-index is also displayed. For $N = 35$, the average computational saving is around a factor 30, which comes down to the difference between one month and one day.

6.2 Linearized quasi-geostrophic equation on the Atlantic Ocean

Let $\Omega \subset \mathbb{R}^2$ be open, bounded and with Lipschitz boundary. The scalar solution fields v, ρ on Ω are said to satisfy the steady one-layer quasi-geostrophic equation in streamline-vorticity formulation if they solve, see [Cavallini and Crisciani, 2013], the equations

$$\begin{cases} \rho = \Delta v & \text{in } \Omega, \\ \mu_3 \mathcal{F}(v,\rho) + \dfrac{\partial v}{\partial x_1} + \mu_1 \rho - \mu_2 \Delta \rho = u & \text{in } \Omega, \\ v = 0 & \text{on } \partial\Omega, \\ \rho = 0 & \text{on } \partial\Omega, \end{cases}$$

where $\mu = (\mu_1, \mu_2, \mu_3) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_{\geq 0}$ are physical parameters representing fluid properties, u represents wind stress and $\mathcal{F}$ describes a nonlinearity given by

$$\mathcal{F}(v,\rho) = \frac{\partial v}{\partial x_1}\frac{\partial \rho}{\partial x_2} - \frac{\partial v}{\partial x_2}\frac{\partial \rho}{\partial x_1},$$

for all v, ρ in a suitable function space specified later. In practice, the parameters μ_1 and μ_3 are smaller than unity. If $\boldsymbol{v}$ denotes the velocity field of a fluid, then $\boldsymbol{v}$ can be recovered from v as $\boldsymbol{v} = (-\frac{\partial v}{\partial x_2}, \frac{\partial v}{\partial x_1})$. As parameter space we take $\mathbb{M} = [10^{-4}, 1] \times [0.07^3, 1] \times [10^{-4}, 0.045^2]$.

We shall work with the corresponding weak form:

$$\begin{aligned} \int_\Omega \rho\tilde{\rho}\,\mathrm{d}x + \int_\Omega \nabla v \cdot \nabla\tilde{\rho}\,\mathrm{d}x &= 0, \\ \mu_3 \int_\Omega \mathcal{F}(v,\rho)\tilde{v}\,\mathrm{d}x + \int_\Omega \frac{\partial v}{\partial x_1}\tilde{v}\,\mathrm{d}x + \mu_1 \int_\Omega \rho\tilde{v}\,\mathrm{d}x + \mu_2 \int_\Omega \nabla\rho\cdot\nabla\tilde{v}\,\mathrm{d}x &= 0, \end{aligned}$$

where $v, \rho, \tilde{v}, \tilde{\rho} \in H_0^1(\Omega)$. Defining $Y = H_0^1(\Omega) \times H_0^1(\Omega)$, $P = Y$ and $U = L^2(\Omega)$, this can be written as

$$A(\mu)y + B(\mu)u = 0, \tag{6.3}$$

where, for all $\mu \in \mathbb{M}$, $A(\mu) : Y \to P^*$ and $B(\mu) \in \mathcal{B}(U, P^*)$ are given by

$$\begin{aligned} \langle A(\mu)(v,\rho), (\tilde{v},\tilde{\rho})\rangle_{P^*P} &= \langle A_0(\mu)(v,\rho), (\tilde{v},\tilde{\rho})\rangle_{P^*P} + \mu_3 \int_\Omega \mathcal{F}(v,\rho)\tilde{v}\,\mathrm{d}x, \\ \langle A_0(\mu)(v,\rho), (\tilde{v},\tilde{\rho})\rangle_{P^*P} &= \int_\Omega \frac{\partial v}{\partial x_1}\tilde{v}\,\mathrm{d}x + \mu_2 \int_\Omega \nabla\rho\cdot\nabla\tilde{v}\,\mathrm{d}x + \mu_1 \int_\Omega \rho\tilde{v}\,\mathrm{d}x \\ &\quad + \int_\Omega \rho\tilde{\rho}\,\mathrm{d}x + \int_\Omega \nabla v\cdot\nabla\tilde{\rho}\,\mathrm{d}x, \\ \langle B(\mu)u, (\tilde{v},\tilde{\rho})\rangle_{P^*P} &= -\int_\Omega u\tilde{v}\,\mathrm{d}x. \end{aligned}$$

Using the Cauchy-Schwarz inequality it is easy to see that $B(\mu)$ is bounded with $\|B(\mu)\|_{\mathcal{B}(U,P^*)} \leq 1$, and that $A_0(\mu)$ is linear and also bounded with $\|A_0(\mu)\|_{\mathcal{B}(Y,P^*)} \leq (3 + |\mu_1| + |\mu_2|)$.

Remark. For the state equation to be almost surely well-posed it is required that $A(\mu)$ is invertible for $\mathbb{P}^{\boldsymbol{\mu}}$-almost every $\mu \in \mathbb{M}$. It is shown in [Barcilon et al., 1988] that $A(\mu)$ is invertible if μ_1 is small enough in comparison to μ_2 and if one instead takes $Y = P = \big(H^2(\Omega) \cap H_0^1(\Omega)\big) \times \big(H^2(\Omega) \cap H_0^1(\Omega)\big)$. Furthermore, [Barcilon et al., 1988] also requires $u \in L^\infty(\Omega)$, leading to a nonreflexive control space. Due to the numerical success obtained with (6.3) (and spaces defined therein) in [Strazzullo et al., 2017], we shall continue to use (6.3) in this work. Stabilization procedures could also be used, such as in [Kim et al., 2015].

[5]The speedup-index is machine dependent. The results are obtained with 6GB of RAM and a 2.60 GHz i5-3230M CPU.

Before considering an OCP governed by the nonlinear quasi-geostrophic equation, we consider a linearized version by taking $\mu_3 = 0$. Therefore, $A(\mu) = A_0(\mu) \in \mathcal{B}(Y, P^*)$. This version of the quasi-geostrophic equation is also known as the linear Stommel-Munk model, [Cavallini and Crisciani, 2013, Kim et al., 2015].

Problem formulation As physical domain Ω we take the one from [Strazzullo et al., 2017]: the part of the Atlantic Ocean between the coasts of Florida North-Africa and Southern Europe. The authors have kindly shared a scaled model of this domain, which we indicate with Ω, as well as a fine mesh $\mathcal{T}_h$ on Ω. The mesh size is denote by h. In Figure 6.6, the mesh is shown.

Let $v_d \in Z$ represent observed data, with observation space $Z = L^2(\Omega) \times L^2(\Omega)$. Given fluid properties μ, we are interested in finding the action of the wind u that, according to the quasi-geostrophic model, would have generated v_d, or at least a state that is close to v_d in the $L^2(\Omega)$ sense. This can be done by solving the following minimization problem:

$$\begin{aligned}
&\text{minimize } J(v, \rho, u) = \frac{1}{2}\int_\Omega |v - v_d|^2 \,\mathrm{d}x + \frac{\alpha}{2}\int_\Omega |u|^2 \,\mathrm{d}x \\
&\text{over all } (v, \rho) \in Y, u \in U, \\
&\text{such that } A(\mu)(v, \rho) + B(\mu)u = 0.
\end{aligned}$$

The constant α is given a small but nonzero value 10^{-5} for the problem to have at most one solution. Notice that in the formulation of Definition 2.2, $C \in \mathcal{B}(Y, Z)$ is the injection operator. Furthermore,

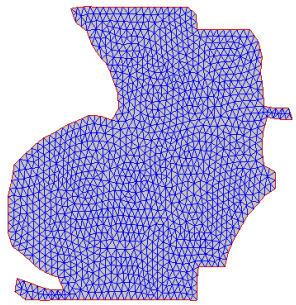

Figure 6.6: The physical domain Ω including a triangulation on this domain.

$M(\mu) \in \mathcal{B}(Z, Z^*)$ and $L(\mu) \in \mathcal{B}(U, U^*)$ are given by

$$\begin{aligned}
\langle M(\mu)z, \tilde{z}\rangle_{Z^*Z} &= \int_\Omega z_1\tilde{z}_1 \,\mathrm{d}x && z = (z_1, z_2), \tilde{z} = (\tilde{z}_1, \tilde{z}_2) \in Z, \\
\langle L(\mu)u, \tilde{u}\rangle_{U^*U} &= \alpha\int_\Omega u\tilde{u} \,\mathrm{d}x && u, \tilde{u} \in U,
\end{aligned}$$

and $g = 0$.

Truth approximation Taking the Finite Elements X_h^1 of (6.2), we set $Y^{\mathcal{N}_Y} = (X_h^1 \times X_h^1) \cap Y$, $U^{\mathcal{N}_U} = X_h^1$ and $P^{\mathcal{N}_P} = Y^{\mathcal{N}_Y}$. As is done in [Strazzullo et al., 2017], we simply assume that $A^{\mathcal{N}}(\mu)$ is invertible and thus the OCP is well-posed for $\mathbb{P}^{\boldsymbol{\mu}}$-a.e. $\mu \in \mathbb{M}$.

Reduced Order Model We build a ROM by using a modification of the partitioned POD approach proposed in Subsection 5.2. Writing $p = (w, q)$ for the adjoint solution, we perform five POD compressions, one for each of the v, ρ, u, w and q components, and aggregate the spaces of state and adjoint. This leaves us with as system of size $9N \times 9N$ to be solved. We then also construct ROMs by leaving out this aggregation step. In that case, we need only solve a system of size $5N \times 5N$. While we do not show that the reduced system is well-posed, notice that an efficient offline phase is ensured through the affine decomposition of Assumption 5.1, as A is affine with $Q_A = 3$ terms, and the terms L, M, B, g and v_d are parameter independent.

Results For a numerical implementation, the desired state v_d is simulated by solving the quasi-geostrophic equation (6.3) with $\mu_1 = \mu_3 = 0$, $\mu_2 = 0.07^3$ and forcing term u given by $(x_1, x_2) \mapsto -\sin(\pi x_2)$. Solving the truth problem for $\mu = (10^{-4}, 0.07^3, 0)$, we obtain the truth approximation $(v^{\mathcal{N}}, \rho^{\mathcal{N}}, u^{\mathcal{N}}, w^{\mathcal{N}}, q^{\mathcal{N}})$. The dimension of the truth problem is 5813×5813.

Afterwards, ROMs are built with $N = 1, \ldots, 20$ using Monte Carlo sampling to get a training set of size 100. Performing Galerkin projection of the truth solution onto the reduced basis space corresponding to $N = 20$, a reduced basis approximation $(v_N, \rho_N, u_N, w_N, q_N)$ is obtained. The desired state v_d, solution component $v^{\mathcal{N}}$ and pointwise difference $v_N - v^{\mathcal{N}}$ for the parameter value $\mu = (10^{-4}, 0.07^3, 0)$ are shown in Figure 6.7.

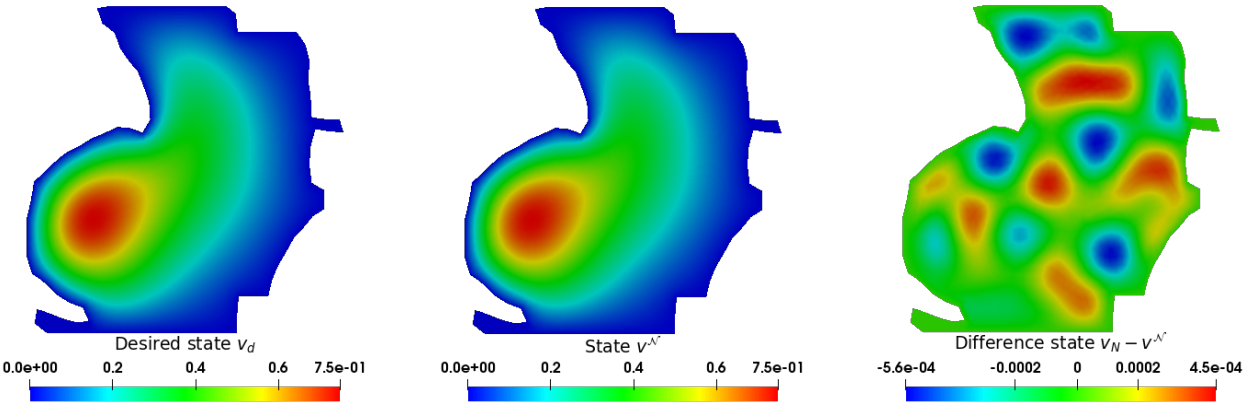

Figure 6.7: Atlantic, linear case: ROM obtained with Monte Carlo sampling from uniform distribution: desired state v_d (*left*), truth state solution component $v^{\mathcal{N}}$ (*center*), and pointwise difference $v_N - v^{\mathcal{N}}$, with $N = 10$ (*right*). The parameter value taken is $\mu = (10^{-4}, 0.07^3, 0)$.

Next, we apply the ROMs constructed with the aggregation approach to compute reduced basis approximations for parameter values in a testing set of size 100. A plot of the average logarithmic errors over this testing set are shown in Figure 6.8. The optimal number of basis functions for the control space is larger than that for the other spaces. Still, a low number of basis vectors is required to ensure small relative errors. We compare these results with ROMs obtained with different quadrature rules in Figure 6.9. Each ROM is constructed twice, once with and once without the aggregation step. All four rules generate a training set of size 100, and only the error for the v-component of the state is shown. The Monte Carlo, Pseudo Random and Gaussian quadrature perform similarly, and notably much better than the Clenshaw-Curtis quadrature rule that is implemented. Furthermore, we note that the ROMs constructed by skipping the aggregation step, actually reach a higher accuracy. While they also require a larger number N of basis functions to be retained, the fact that aggregation is skipped means that only a system of dimension $5N \times 5N$ instead of $9N \times 9N$ must be solved online. For example, the most accurate ROM constructed without the aggregation approach is obtained with the Pseudo-Random Sampler with $N = 19$, which leads to a system to be solved online of size 95×95. The optimal ROM with an aggregation approach is constructed with the Monte Carlo sampler with $N = 11$, so that the online system is of size 99×99. Furthermore, the optimal ROM constructed without the aggregation approach, is on average more accurate by about two orders of magnitude.

In Figure 6.10 (*left*) and Figure 6.11 (*left*) also ROMs for a Beta$(75, 75)$, Beta$(5, 1)$ and Loguniform distribution are considered. Each is constructed by aggregating state and adjoint. For the Beta(75,75) distribution, the Pseudo Random sampler picks out an extra useful direction in the solution manifold of the v-component. With only $N = 5$ very high accuracy is obtained. For the Beta(5,1) distribution, that puts more probability mass on the larger parameter values, the Monte Carlo and Pseudo Random samplers seem to be preferable over Gaussian quadrature. For both distributions, the difference between the quadratures is small. Interestingly, the same can not be concluded in the Loguniform case, which emphasizes small parameter values, as the performance of both the Clenshaw-Curtis and the Gaussian quadrature is poor, and their accuracies vary more extensively over the testing set.

The same distributions are considered in Figures 6.10 (*right*) and 6.11 (*right*), but this time the cor-

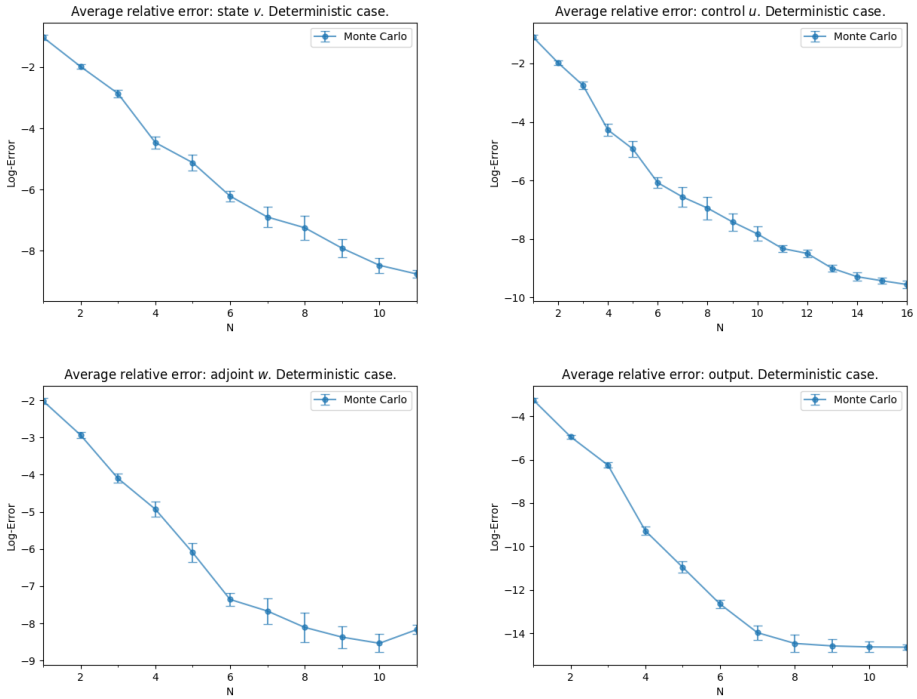

Figure 6.8: Atlantic, linear case: average logarithmic relative errors for the v-component of state (*top left*), control (*top right*), w-component of adjoint (*bottom left*) and output (*bottom right*) in the deterministic case, as a function of N. ROM obtained via Monte Carlo sampling.

responding ROMs are constructed skipping the aggregation step. Let us compare the optimal ROMs contructed with and without the aggregation procedure. For the Beta$(75, 75)$ distribution, the optimal ROM with aggregation is constructed with a Pseudo Random sampler for $N = 6$, and without aggregation with the Monte Carlo sampler for $N = 8$. For the Beta$(5, 1)$ distribution, these are respectively the Monte Carlo sampler with $N = 5$ and the Monte Carlo sampler with $N = 10$. In both cases, the accuracy of the ROM obtained without aggregation is at least as high as for the ROM obtained with aggregation.

For the Loguniform distribution, the optimal ROMs are the Monte Carlo rules with $N = 35$ and Pseudo Random sampler with $N = 40$, but we do note that in this case we stopped the simulations at $N = 40$ so that higher accuracy can possibly be obtained.

Note that the errors corresponding to the aggregation approach do decay more monotonously. Non monotonous decay can occur in general, due to the fact that the Galerkin Projection is not an orthogonal projection but a skewed projection.

Finally, we present the speedup-index in the deterministic case with Monte Carlo sampling in Table 6.2. With $N = 12$, an average speedup of over 70 times is achieved using a ROM constructed with the aggregation step. This speedup is obtained with a ROM of $N = 20$ in case the aggregation step is skipped. Notice, however, that the deviations in the speedup index are larger for ROMs constructed without the aggregation step.

6.3 Nonlinear quasi-geostrophic equation on the Atlantic Ocean

Having investigated the performance of a ROM for the linearized version of the quasi-geostrophic state equation, we now relax the condition $\mu_3 = 0$ and thus allow the nonlinearity to enter the system. Apart from this modification, the control problem is the same as in Subsection 6.2.

Optimal Control for problems with a nonlinear state equation By Theorem 3.4, the OCP for μ in a set of probability one can be solved by solving the system (3.2) for that μ, if $(DA(\mu))y \in \mathcal{B}(Y, P^*)$

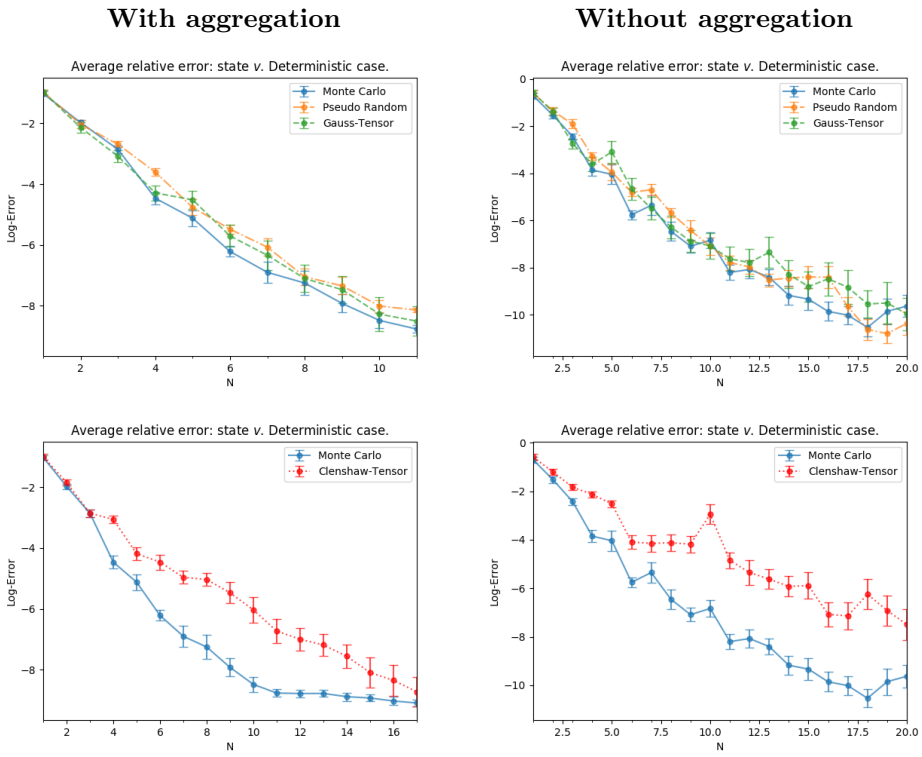

Figure 6.9: Atlantic, linear case: average logarithmic relative errors in the v-component in the deterministic case, for ROMs constructed with aggregation (*left*) and without aggregation (*right*), as a function of N. Various quadrature rules are compared.

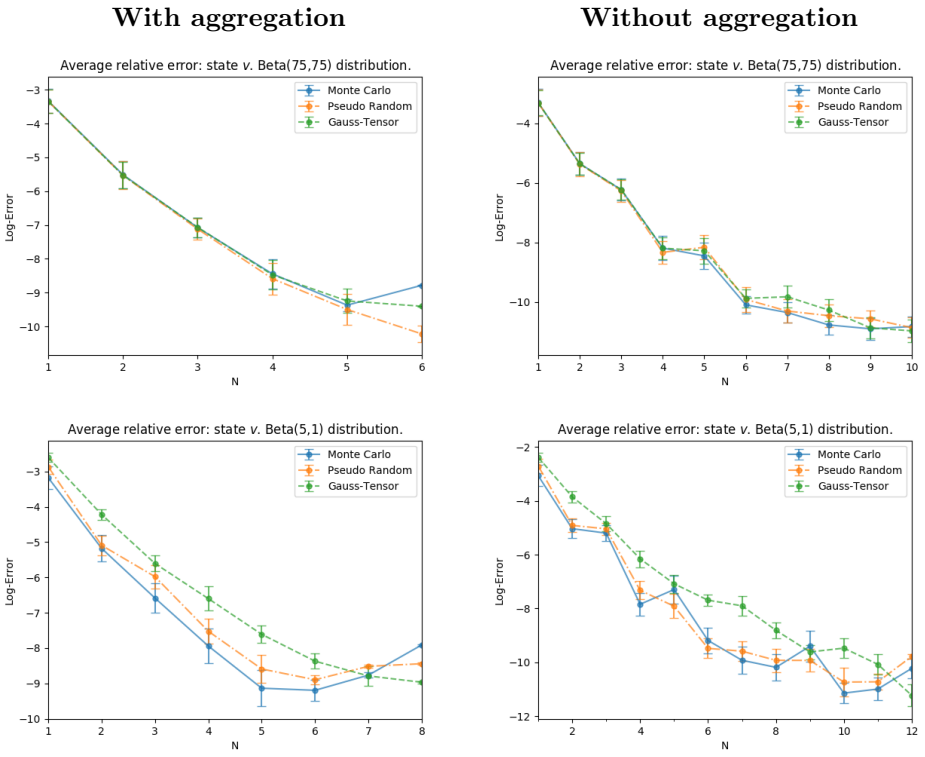

Figure 6.10: Atlantic, linear case: average logarithmic relative errors in the v-component under a Beta$(75, 75)$ distribution (*top*) and Beta$(5, 1)$ distribution (*bottom*), for ROMs constructed with aggregation (*left*) and without aggregation (*right*), as a function of N. Various quadrature rules are compared.

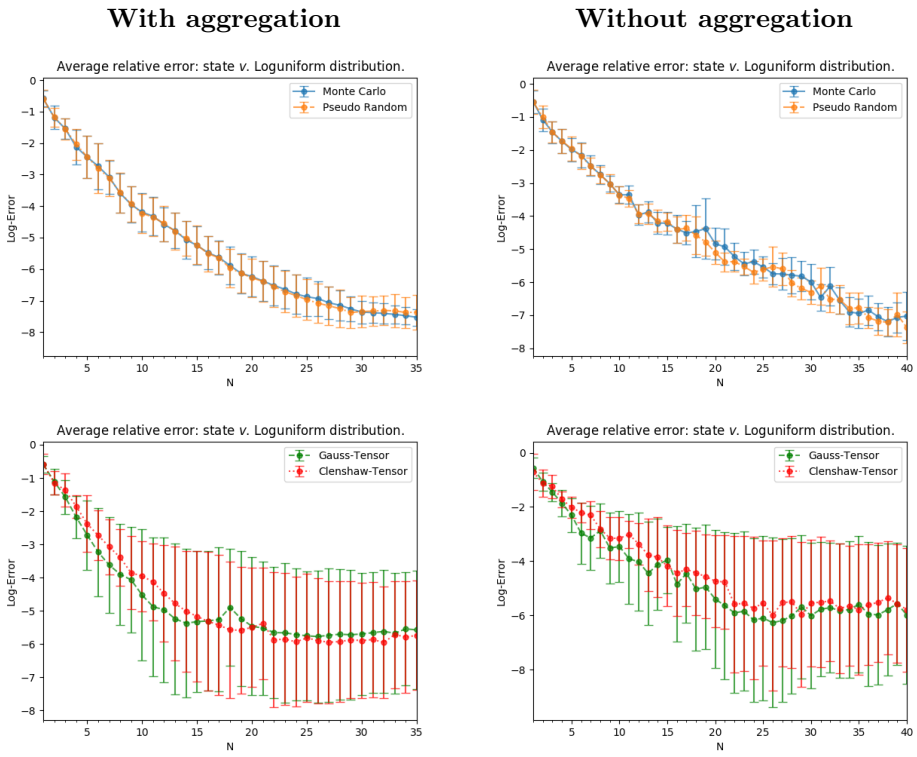

Figure 6.11: Atlantic, linear case: average logarithmic relative errors in the v-component under a Loguniform distribution, for ROMs constructed with aggregation (*left*) and without aggregation (*right*), as a function of N. Various quadrature rules are compared.

With aggregation

N	average	min	max	deviation
2	99.6	55.9	129.9	9.9
4	95.5	63.6	123.0	8.9
8	84.9	61.8	110.7	7.9
12	72.8	54.7	94.6	6.3
16	58.3	41.4	75.4	5.7
20	45.7	33.0	58.9	4.3

Without aggregation

N	average	min	max	deviation
2	99.8	51.4	145.9	17.1
4	98.4	60.8	144.9	14.8
8	96.0	57.6	128.4	14.2
12	83.9	48.2	125.1	14.3
16	80.5	48.4	114.5	12.3
20	72.2	37.6	103.0	12.1

Table 6.2: Atlantic, linear case: sample average, minimal value, maximal value, and sample standard deviation of speedup-index obtained with a testing set of size 100. Deterministic case. ROM obtained by Monte Carlo sampling, with aggregation (*left*) and without aggregation (*right*).

has a bounded inverse for every $y \in Y$, but we do not dwell on this assumption to hold. Writing $y = (v, \rho)$ and taking $(\tilde{v}, \tilde{\rho}) \in Y$, $p = (w, q) \in P$, we have

$$\begin{aligned}\langle((DA(\mu))y)(\tilde{v},\tilde{\rho}),p\rangle_{P^*P} &= \langle A_0(\mu)(\tilde{v},\tilde{\rho}),(w,q)\rangle_{P^*P} \\ &\quad + \mu_3\int_\Omega \mathcal{F}(\tilde{v},\rho)w\,\mathrm{d}x + \mu_3\int_\Omega \mathcal{F}(v,\tilde{\rho})w\,\mathrm{d}x\end{aligned}$$

If we denote the adjoint variable by $p(\mu) = (w(\mu), q(\mu))$, the adjoint equation, the second equation of (3.2), in this case reads (suppressing the injection $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ in the last term)

$$\begin{aligned}0 &= \langle A_0(\mu)(\tilde{v},\tilde{\rho}),(w(\mu),q(\mu))\rangle_{P^*P} - \mu_3\int_\Omega \mathcal{F}(v(\mu),w(\mu))\tilde{\rho}\,\mathrm{d}x \\ &\quad - \mu_3\int_\Omega \mathcal{F}(w(\mu),\rho(\mu))\tilde{v}\,\mathrm{d}x + \langle M(\mu)(v(\mu)-v_d,\rho(\mu)),(\tilde{v},\tilde{\rho})\rangle_{Z^*Z},\end{aligned} \tag{6.4}$$

while the state equation and optimality equation are still of the same form as those in (3.3).

Truth formulation and Reduced Order Model The truth formulation for this nonlinear problem is obtained no differently than in the linear case. The spaces $Y^{\mathcal{N}_Y}, U^{\mathcal{N}_U}$ and $P^{\mathcal{N}_P}$ are as in the linear case. The truth formulation then is formed by solving (6.4) together with the last two equations of (3.4), with Y, U, P replaced by $Y^{\mathcal{N}_Y}, U^{\mathcal{N}_U}, P^{\mathcal{N}_P}$. Being a nonlinear system, it cannot be solved at once. An iterative procedure can be used to find a solution, and we choose to employ Newton's method to solve the truth problem. While we do not theoretically show well-posedness of the truth problem, we decide to remain in a *mild nonlinear setting* by the choice of parameter space, see [Cavallini and Crisciani, 2013]. This has always led to a convergent Newton solver. For higher values of the nonlinear parameter μ_3, one shall need to stabilize the system at hand.
A ROM is constructed using the same methods as for the linear case. Again, well-posedness is assumed and a Newton iteration procedure is used to solve the reduced system.

Results This time the desired state v_d is simulated by solving the quasi geostrophic equation (6.3) with $\mu = (0, 0.07^3, 0.07^2)$ and forcing term $(x_1, x_2) \mapsto -\sin(\pi x_2)$. Let us first assume the deterministic setting. The truth problem is solved for the parameter value $\mu = (10^{-4}, 0.07^3, 0.045^2)$.

Assuming $\boldsymbol{\mu}$ is uniformly distributed, we construct a ROM for $N = 1, \dots, 15$, with the aggregation approach. For $N = 15$ a reduced solution component v_N was computed for $\mu = (10^{-4}, 0.07^3, 0.045^2)$. The desired state v_d, truth solution component $v^{\mathcal{N}}$ and pointwise difference $v_N - v^{\mathcal{N}}$ for this parameter value are shown in Figure 6.12. Furthermore, the objective functional is approximated with an accuracy of 13 digits. The average logarithmic error of the output and v-component are displayed in Figure

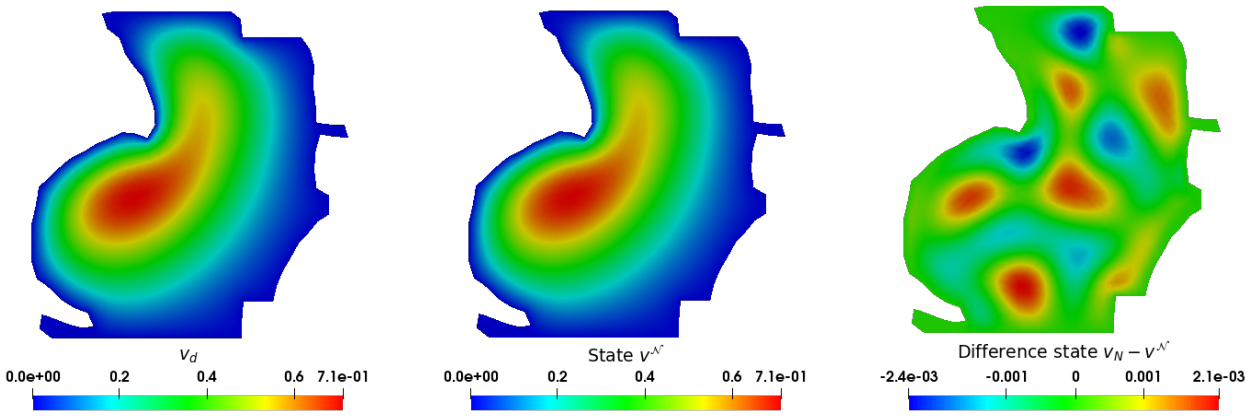

Figure 6.12: Atlantic, nonlinear case: ROM obtained with Monte Carlo sampling from uniform distribution desired state v_d (*left*), truth solution component $v^{\mathcal{N}}$ (*center*), and pointwise difference $v_N - v^{\mathcal{N}}$ (*right*), with $N = 10$. The parameter value taken is $\mu = (10^{-4}, 0.07^3, 0.045^2)$.

6.13 (*top left*). Once more exponential decay can be observed. In Figure 6.13 (*center left, bottom left*) a Beta(75,75) distribution and a Loguniform distribution on each of the three parameter domains is considered. In each case, different quadrature rules are compared, and ROMs are constructed with the aggregation approach. The Beta(75,75) distribution again reduces the complexity of the model, which results in needing only $N = 5$ to reconstruct the discrete solution manifold well. The Loguniform distribution achieves similar accuracy after $N = 25$. It should be noted that the Monte Carlo and Pseudo Random rules taken use a training set of size 100, while the Gaussian and Clenshaw-Curtis rules use a training set of size 125.

Despite working in the mild nonlinear setting, for some parameter value the Newton Iteration procedure has diverged for the ROM with $N = 24$ and a Gauss sampler. This results in the large deviation that is observed.

The same ROMs are constructed leaving out the aggregation step, and the corresponding results are shown in Figure 6.13 (*right*). As in the example of Subsection 6.2, we can conclude that without the need of the aggregation step, the obtained results are of an accuracy that is at least as high as in an aggregation approach, while saving online computation time in general.

In Table 6.3 the speedup-index for the Monte Carlo Sampler in the deterministic setting is shown. The speedups are small due to the nonlinearity, as the system that needs to be solved online still needs

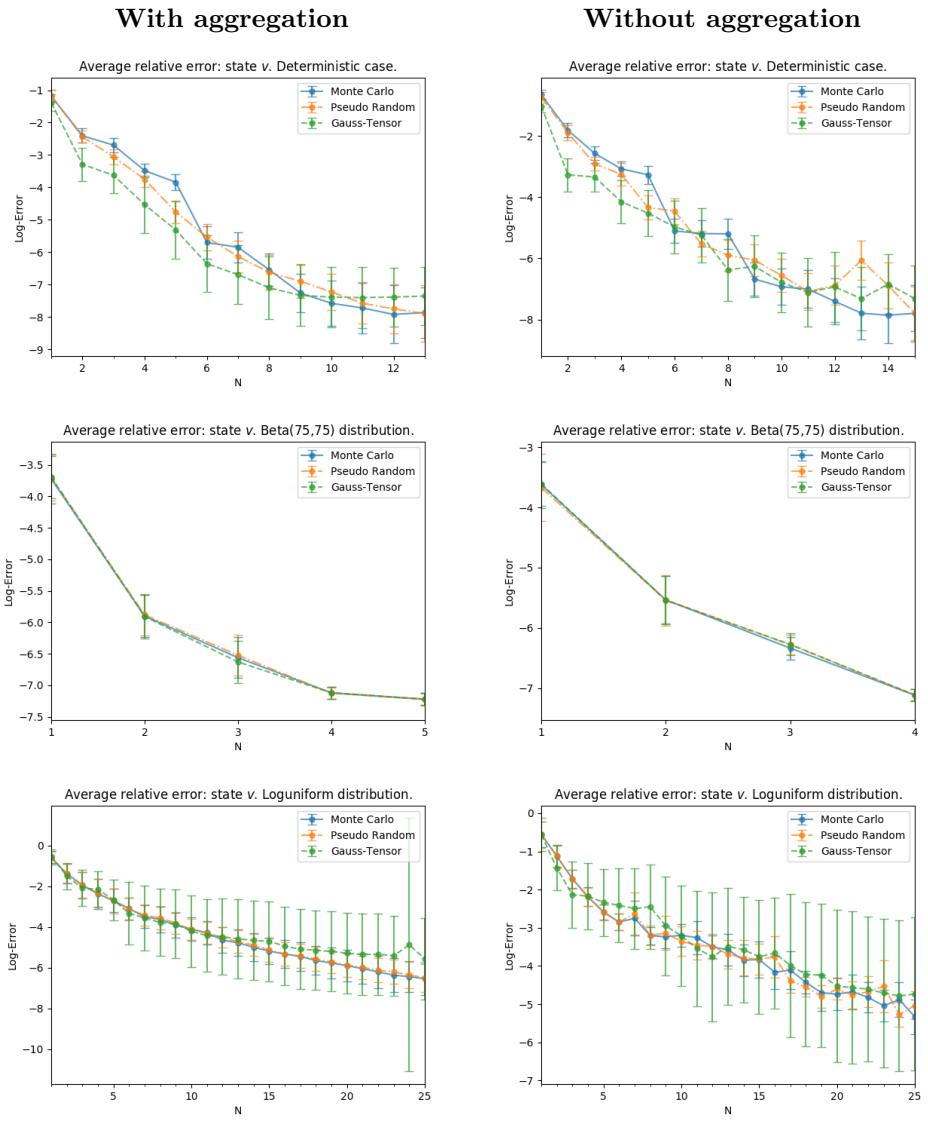

Figure 6.13: Atlantic, nonlinear case: average logarithmic relative errors for the state v-component in the deterministic case (*top*), under a Beta$(75, 75)$ distribution (*center*) and under a Loguniform distribution (*bottom*), for ROMs constructed with aggregation (*left*) and without aggregation (*right*), as a function of N. Various quadrature rules are compared.

to be assembled in a number of computations that depends on $\mathcal{N}$. Furthermore, the Newton iteration used to solve this system can take many steps to converge. While the ROMs are still effective, this lack of efficiency renders the ROM less useful in this application. Nevertheless, there are strategies which can overcome this issue, the interested reader may refer to see [Barrault et al., 2004].

With aggregation

N	average	min	max	deviation
2	1.8	1.3	2.3	0.20
3	1.7	1.3	2.1	0.20
6	1.3	1.0	1.7	0.18
9	1.0	0.80	1.4	0.14
12	0.83	0.60	1.1	0.12
15	0.66	0.45	0.86	0.097

Without aggregation

N	average	min	max	deviation
2	1.8	1.4	2.9	0.25
3	1.8	1.1	3.1	0.27
6	1.6	1.0	2.6	0.26
9	1.4	0.9	2.4	0.25
12	1.3	0.83	2.0	0.20
15	1.1	0.75	1.9	0.18

Table 6.3: Atlantic, nonlinear case: sample average, minimal value, maximal value, and sample standard deviation of speedup-index obtained with a testing set of size 100. Deterministic case. ROM obtained by Monte Carlo sampling, with aggregation (*left*) and without aggregation (*right*).

7 Conclusions

As an alternative to the Lagrangian approach applicable to linear-quadratic Optimal Control Problems with full-admissibility, we have studied OCPs which may have admissibility constraints or which are not of linear-quadratic nature by using the adjoint approach. As a by-product we have concluded that to establish well-posedness of OCPs and their approximations, the use of aggregate spaces can be avoided, because the fundamental requirement is invertibility of the reduced state operator. Having said this, in the coercive case the use of aggregate spaces is still useful, because it guarantees this invertibility.
We have also studied OCPs with random inputs, and used the weighted POD algorithm to construct ROMs for OCPs with full admissibility. The weighted POD has enabled us to incorporate information on parameter distributions, that originate from, for example, uncertainties involved in experimental measurements. The ROMs then allowed us to accurately and efficiently compute approximations to linear-quadratic OCPs with full admissibility in several marine science scenarios. We have also embedded a scenario in which the governing equation is the nonlinear single-layer steady quasi-geostrophic equation in an Uncertainty Quantification context.
In our scenarios for which the governing equation was not elliptic, we have numerically confirmed that it is not needed to use aggregate spaces. Indeed, the ROMs that were constructed without this aggregation performed at least as well, and in most cases resulted in a speedup through a smaller system that has to be solved in the online phase. Furthermore, we have considered various types of distributions on the parameter space. In each case, the constructed ROMs effectively incorporated this information. As the weighted POD algorithm depends on a quadrature rule, we have explored several implementations of quadrature rules. No numerical evidence that one specific rule should be preferred over the others has been found, although the chosen implementation of the Clenshaw-Curtis rule should be avoided. Henceforth, one might as well use the Monte Carlo sampler, as it is the simplest one.

Acknowledgements

We acknowledge the funding granted via the European Eramus+ project and the support by European Union Funding for Research and Innovation – Horizon 2020 Program – in the framework of European Research Council Executive Agency: Consolidator Grant H2020 ERC CoG 2015 AROMA-CFD project 681447 "Advanced Reduced Order Methods with Applications in Computational Fluid Dynamics". We also acknowledge the INDAM-GNCS project "Advanced intrusive and non-intrusive model order reduction techniques and applications" and the PRIN 2017 "Numerical Analysis for Full and Reduced Order Methods for the efficient and accurate solution of complex systems governed by Partial Differential Equations" (NA-FROM-PDEs). The computations in this work have been performed with RBniCS [rbn, 2015] library, developed at SISSA mathLab, which is an implementation in FEniCS [Logg et al., 2012] of several reduced order modelling techniques; we acknowledge developers and contributors to both libraries.

References

[rbn, 2015] (2015). RBniCS - reduced order modelling in FEniCS. `https://www.rbnicsproject.org/`.

[Bader et al., 2017] Bader, E., Grepl, M. A., and Veroy, K. (2017). A Certified Reduced Basis Approach for Parametrized Optimal Control Problems with Two-Sided Control Constraints. In *Model Reduction of Parametrized System*, pages 37–54. Springer International Publishing, Cham.

[Ballarin et al., 2015] Ballarin, F., Manzoni, A., Quarteroni, A., and Rozza, G. (2015). Supremizer stabilization of POD–Galerkin approximation of parametrized steady incompressible Navier–Stokes equations. *International Journal for Numerical Methods in Engineering*, 102(5):1136–1161.

[Barcilon et al., 1988] Barcilon, V., Constantin, P., and Titi, E. (1988). Existence of Solutions to the Stommel-Charney Model of the Gulf Stream. *SIAM Journal on Mathematical Analysis*, 19(6):1355–1364.

[Barrault et al., 2004] Barrault, M., Maday, Y., Nguyen, N., and Patera, A. T. (2004). An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672.

[Burkardt et al., 2006] Burkardt, J., Gunzburger, M. D., and Lee, H.-C. (2006). POD and CVT-based reduced-order modeling of Navier-Stokes flows. *Computer Methods in Applied Mechanics and Engineering*, 196(1):337–355.

[Carere, 2020] Carere, G. (2020). Reduced Order Methods for Optimal Control Problems constrained by PDEs with random inputs and applications. Master's thesis, University of Amsterdam and SISSA.

[Cavallini and Crisciani, 2013] Cavallini, F. and Crisciani, F. (2013). *Quasi-Geostrophic Theory of Oceans and Atmosphere: Topics in the Dynamics and Thermodynamics of the Fluid Earth*, volume 45. Springer Science & Business Media, New York.

[Chapelle et al., 2013] Chapelle, D., Gariah, A., Moireau, P., and Sainte-Marie, J. (2013). A Galerkin strategy with Proper Orthogonal Decomposition for parameter-dependent problems – Analysis, assessments and applications to parameter estimation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(6):1821–1843.

[Chen et al., 2017] Chen, P., Quarteroni, A., and Rozza, G. (2017). Reduced Basis Methods for Uncertainty Quantification. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):813–869.

[Dedè, 2010] Dedè, L. (2010). Reduced Basis Method and A Posteriori Error Estimation for Parametrized Linear-Quadratic Optimal Control Problems. *SIAM Journal on Scientific Computing*, 32(2):997–1019.

[Griebel and Harbrecht, 2018] Griebel, M. and Harbrecht, H. (2018). Singular value decomposition versus sparse grids: refined complexity estimates. *IMA Journal of Numerical Analysis*, 39(4):1652–1671.

[Haasdonk, 2017] Haasdonk, B. (2017). Reduced Basis Methods for Parametrized PDEs–A Tutorial Introduction for Stationary and Instationary Problems. In *Model Reduction and Approximation*, volume 15, chapter 2, pages 65–136. SIAM Publications, Philadelphia.

[Hesthaven et al., 2016] Hesthaven, J., Rozza, G., and Stamm, B. (2016). *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer International Publishing, Milano.

[Hinze et al., 2009] Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S. (2009). *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer Netherlands.

[Kärcher and Grepl, 2014] Kärcher, M. and Grepl, M. A. (2014). A certified reduced basis method for parametrized elliptic optimal control problems. *ESIAM:COCV*, 20(2):416–441.

[Kärcher et al., 2018] Kärcher, M., Tokoutsi, Z., Grepl, M. A., and Veroy, K. (2018). Certified Reduced Basis Methods for Parametrized Elliptic Optimal Control Problems with Distributed Controls. *Journal of Scientific Computing*, 75(1):276–307.

[Kim et al., 2015] Kim, T., Iliescu, T., and Fried, E. (2015). B-spline based finite-element method for the stationary quasi-geostrophic equations of the ocean. *Computer Methods in Applied Mechanics and Engineering*, 286:168–191.

[Lions, 1971] Lions, J. (1971). *Optimal control of systems governed by partial differential equations*, volume 170 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin Heidelberg.

[Logg et al., 2012] Logg, A., Mardal, K., and Wells, G. (Berlin, 2012). *Automated Solution of Differential Equations by the Finite Element Method*, volume 84. Springer-Verlag.

[Negri et al., 2015] Negri, F., Manzoni, A., and Rozza, G. (2015). Reduced basis approximation of parametrized optimal flow control problems for the Stokes equations. *Computers & Mathematics with Applications*, 69(4):319–336.

[Negri et al., 2013] Negri, F., Rozza, G., Manzoni, A., and Quarteroni, A. (2013). Reduced Basis Method For Parametrized Elliptic Optimal Control Problems. *SIAM Journal on Scientific Computing*, 35(5):A2316–A2340.

[Prud'homme et al., 2001] Prud'homme, C., Veroy, D., Machiels, L., Maday, Y., Patera, A. T., and Turinici, G. (2001). Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *ASME. Journal of Fluids Engineering*, 124(1):70–80.

[Quarteroni and Valli, 1994] Quarteroni, A. and Valli, A. (1994). *Numerical Approximation of Partial Differential Equations*, volume 23 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg.

[Rozza et al., 2013] Rozza, G., Huynh, D., and Manzoni, A. (2013). Reduced basis approximation and a posteriori error estimation for Stokes flows in parametrized geometries: roles of the inf-sup stability constants. *Numerische Mathematik*, 125:115–152.

[Rozza et al., 2008] Rozza, G., Huynh, D. B. P., and Patera, A. T. (2008). Reduced Basis Approximation and a Posteriori Error Estimation for Affinely Parametrized Elliptic Coercive Partial Differential Equations. *Archives of Computational Methods in Engineering*, 15(3):229–275.

[Rozza and Veroy, 2007] Rozza, G. and Veroy, K. (2007). On the stability of the reduced basis method for Stokes equations in parametrized domains. *Computer Methods in Applied Mechanics and Engineering*, 196(7):1244–1260.

[Schwab and Todor, 2006] Schwab, C. and Todor, R. (2006). Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122.

[Strazzullo, 2016] Strazzullo, M. (2015-2016). Reduced Order Methods for Parametrized Optimal Flow Control Problems. Master's thesis, University of Trieste and SISSA.

[Strazzullo et al., 2017] Strazzullo, M., Ballarin, F., Mosetti, R., and Rozza, G. (2017). Model Reduction for Parametrized Optimal Control Problems in Environmental Marine Sciences and Engineering. *SIAM Journal on Scientific Computing*, 40(4):B1055–B1079.

[Sullivan, 2015] Sullivan, T. (2015). *Introduction to Uncertainty Quantification*, volume 63 of *Texts in Applied Mathematics*. Springer International Publishing.

[Torlo et al., 2018] Torlo, D., Ballarin, F., and Rozza, G. (2018). Stabilized Weighted Reduced Basis Methods for Parametrized Advection Dominated Problems with Random Inputs. *SIAM/ASA Journal of Uncertainty Quantification*, 6(4):1475–1502.

[Venturi, 2016] Venturi, L. (2015-2016). Weighted Reduced Order Methods For Parametrized PDES In Uncertainty Quantification Problems. Master's thesis, University of Trieste and SISSA.

[Venturi et al., 2019a] Venturi, L., Ballarin, F., and Rozza, G. (2019a). A Weighted POD Method for Elliptic PDEs with Random Inputs. *Journal of Scientific Computing*, 81(1):136–153.

[Venturi et al., 2019b] Venturi, L., Torlo, D., Ballarin, F., and Rozza, G. (2019b). Weighted Reduced Order Methods for Parametrized Partial Differential Equations with Random Inputs. In *Uncertainty Modeling for Engineering Applications*, pages 27–40. Springer International Publishing, Cham.