A forward approach to numerical data assimilation

Christian Clason* Peter Heppergerger†

Variational data assimilation problems are concerned with computing unknown initial values for the simulation and prediction of natural phenomena, most notably in weather prediction, and are usually solved via an ill-posed optimal control problem for the initial state at the time of the first available measurements. An alternative "forward" approach focuses on computation of the final state after this interval—which is just as suitable for prediction purposes—and is well-posed without additional regularization. Specifically, it is possible to compute projections of the unknown final state on all elements of an orthonormal basis, which theoretically allows for the complete reconstruction of the final state. In this paper, an efficient numerical method for linear evolution equations of diffusive type is presented, and convergence of the numerical approximation based on a discontinuous Galerkin discretization is proved. The key of this method is the computation of an adaptively ordered orthonormal basis using proper orthogonal decomposition. Numerical examples for a scalar convection-diffusion equation in two and three dimensions show the effectiveness of the method.

1 Introduction

Data assimilation problems are concerned with determining the initial condition at $t = 0$ in evolution equations from distributed or boundary measurements over a time interval $[0, T_0]$, for the purpose of calculating the state at later times. Such problems arise, for example, in weather or climate prediction [13], in oceanography [9] and in geophysics [6], and in general lead to ill-posed problems (such as the backward heat equation). The current standard approach (called 4DVAR [15, 25]) uses optimal control

*Institute for Mathematics and Scientific Computing, Karl-Franzens-Universität Graz, 8010 Graz, Austria, (christian.clason@uni-graz.at).

†Zentrum Mathematik, Technische Universität München, 85748 Garching bei München, Germany (hepperger@ma.tum.de).

techniques for minimizing a suitable cost function involving the observations and the state equation, including a regularization method and an optimality system making use of the adjoint state. The state in the time interval $[0, T_1]$, $T_1 > T_0$, is then calculated via the solution of a classical initial boundary value problem. (Other methods in use include Kalman filters [9, 22] and the representer method [3, 5, 4].) The 4DVAR method is now part of most operational weather prediction models, and can be the most time consuming part of a forecast cycle.

An alternative method [19] computes from distributed observations the state at the final time $t = T_0$, which can serve as an initial condition for the state equation in the time interval $[T_0, T_1]$, for which the prediction is desired. This approach can be thought of as a variational analog of forward data assimilation techniques (such as Kalman filters), to which it is, however, not directly related. The "variational forward assimilation" problem is well-posed, as can be shown using an observability estimate for the adjoint state equation derived via a Carleman estimate (cf. [19]). Specifically, it is stable with respect to noise in the data even without regularization. Its solution can be computed by solving a series of control problems for elements of an orthonormal basis of a finite dimensional subspace.

In this paper, we present an efficient numerical method for solving the "variational forward assimilation" problem for a general linear convection-diffusion equation. By using proper orthogonal decomposition to construct an adaptively ordered orthonormal basis and a discontinuous Galerkin discretization in time, the computation can be carried out very efficiently. It is shown that this discrete problem has a unique solution, which converges to the sought final state. The main advantages of the proposed approach are the high stability with respect to measurement errors, and its inherent pre-computability and parallelizability. Furthermore, it is free of regularization parameters which have to be chosen dependent on the data. We demonstrate the effectiveness of this method on examples in two and three dimensions. For the purposes of presentation, we focus here on linear convection-diffusion equations, but the approach is also directly applicable to more complex state equations such as the linearized Navier-Stokes equations.

We now make the above more precise, and then state the projection formula and stability estimates from [19] which are fundamental for this approach. We conclude with an outline of the entire forward data assimilation procedure, which also serves as an overview of the organization of this article.

1.1 Problem formulation

We consider a linear scalar convection-diffusion equation in a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, with Lipschitz boundary Γ, on the time interval $[0, T]$ for a given $T > 0$:

$$\begin{cases} \partial_t y - \nabla \cdot (c \nabla y) + b^T \nabla y = f, & \text{in } \Omega \times (0, T), \\ y = 0, & \text{on } \Gamma \times [0, T], \end{cases} \tag{1}$$

with a diffusion tensor $c \in \mathrm{C}^1(0, T; \mathrm{C}^\infty(\Omega))^{d \times d}$, a flux $b \in \mathrm{C}^1(0, T; \mathrm{C}^\infty(\Omega))^d$, and a right hand side $f \in \mathrm{L}^2(0, T; \mathrm{L}^2(\Omega))$. The entries c_{ij} of the diffusion tensor are supposed to

satisfy an ellipticity condition: There exists a constant $\gamma > 0$ independent of x and t such that

$$\sum_{i,j=1}^{d} c_{ij}(x,t)\xi_i\xi_j \geq \gamma \|\xi\|^2_{\mathbb{R}^d}$$

holds for all $(x,t) \in \Omega \times [0,T]$ and all $\xi \in \mathbb{R}^d$. We refer to [20] for a precise statement of the function space containing solutions of (1) for which no additional initial condition is specified.

Now let ω be a nonempty open subset of Ω with characteristic function χ_ω. Furthermore, let $v \in L^2(0,T;L^2(\Omega))$ and $\varphi_T \in L^2(\Omega)$ be given. We also introduce the controlled adjoint problem of (1):

$$\begin{cases} -\partial_t \varphi - \nabla \cdot (c^T \nabla \varphi) - \nabla \cdot (b\varphi) = v\chi_\omega, & \text{in } \Omega \times (0,T), \\ \varphi = 0, & \text{on } \Gamma \times [0,T], \\ \varphi(\cdot,T) = \varphi_T, & \text{in } \Omega. \end{cases} \tag{2}$$

Thus, $v\chi_\omega(x)$ defines a control acting on ω.

The following proposition summarizes some statements concerning existence and regularity of the weak solution of (2):

Proposition 1.1. *For given $\varphi_T \in L^2(\Omega)$ and $v\chi_\omega \in L^2(0,T;L^2(\omega))$, equation (2) has a unique weak solution, which satisfies*

$$\varphi \in L^2(0,T;H_0^1(\Omega)) \cap C(0,T;L^2(\Omega)), \qquad \partial_t \varphi \in L^2(0,T;H^{-1}(\Omega)). \tag{3}$$

If in addition $\varphi_T \in H^1(\Omega)$ and $\partial_t v\chi_\omega \in L^2(0,T;L^2(\omega))$ holds, this solution satisfies

$$\begin{cases} \varphi \in L^2(0,T;H^2(\Omega)) \cap C(0,T;H^1(\Omega)) & \text{and} \\ \partial_t \varphi \in L^2(0,T;H^1(\Omega)) \cap C(0,T;L^2(\Omega)), \end{cases} \tag{4}$$

as well as the following estimate:

$$\begin{aligned} &\operatorname{ess\,sup}_{0\leq t\leq T} \|\varphi(t)\|_{L^2(\Omega)} + \|\varphi\|_{L^2(0,T;H^1(\Omega))} + \|\partial_t \varphi\|_{L^2(0,T;L^2(\Omega))} \\ &\qquad \leq C\left(\|v\chi_\omega\|_{L^2(0,T;L^2(\omega))} + \|\varphi_T\|_{H^1(\Omega)}\right) \end{aligned} \tag{5}$$

Proof. Existence, uniqueness and regularity property (3) follow from [28, Ths. 26.1, 25.5]. The higher regularity given in (4) is a direct consequence of [18, Th. 6.2] and [28, Ths. 27.2, 25.5]. The estimate (5) can be found combining [8, 5.9, Th. 2] and again [18, Th. 6.2]. □

We consider then the following problem:

Problem 1.2. *Given $y\chi_\omega \in L^2(0,T;L^2(\omega))$, find $y(T) \in L^2(\Omega)$, where y satisfies equation (1).*

In other words, we are looking for a final state $y(T)$ of the convection-diffusion equation, which is consistent with equation (1) and the given measurement $y\chi_\omega$.

1.2 Well-posedness and projection formula

We state the main results from [19], showing uniqueness and stability for the solution of the forward data assimilation problem:

Theorem 1.3 (Théorème 1, [19]). *There exists a constant $C > 0$ depending only on Ω, T and ω such that for a solution y of (1), the following estimate holds:*

$$\|y(T)\|^2_{L^2(\Omega)} \leq C \left(\int_0^T \int_\omega |y|^2 dxdt + \int_0^T \int_\Omega |f|^2 dxdt \right). \tag{6}$$

This theorem follows from a Carleman estimate for second order parabolic equations and standard energy estimates using a cutoff-function to remove the dependence on the unknown initial condition.

It is well known (c.f., e.g., [17]) that an observability estimate such as (6) implies null-controllability of the adjoint equation. Hence, we have:

Corollary 1.4. *Let $\omega \subset \Omega$ be an open subset, and let $T > 0$ and $\varphi_T \in L^2(\Omega)$ be given. Then there exists a control $v = v(\varphi_T) \in L^2(0,T;L^2(\Omega))$, such that the solution φ of the adjoint equation (2) satisfies $\varphi(0) = 0$.*

To fix a unique control, we take v to be the function of minimal $L^2(0,T;L^2(\omega))$ norm among all admissible controls. Proposition 1 in [10] states that there are coefficients $C_\Omega, K_{T,c,b} > 0$ such that this control v satisfies

$$\|v(\varphi_T)\|_{L^2(0,T;L^2(\Omega))} \leq e^{C_\Omega K_{T,c,b}} \|\varphi_T\|_{L^2(\Omega)}. \tag{7}$$

From this null-controllability result, by multiplying the state equation with the null-controlled adjoint state and integrating by parts, we arrive at the following equality for the projection of the unknown final state onto a function in $L^2(\Omega)$:

Corollary 1.5. *If y satisfies (1), then we have for all $\varphi_T \in L^2(\Omega)$ that*

$$\int_\Omega y(T)\varphi_T \, dx = \int_0^T \int_\Omega f\varphi \, dxdt - \int_0^T \int_\omega yv(\varphi_T) \, dxdt, \tag{8}$$

where φ and $v(\varphi_T)$ are the null-controlled solution and corresponding null-control according to Corollary 1.4.

Since only known functions appear on the right hand side of (8), this allows the computation of projections of the final state on any given $\varphi_T \in L^2(\Omega)$.

While it follows immediately from (8) that these projections—and hence any finite dimensional approximation of $y(T)$ computed this way—are stable with respect to perturbations of $y\chi_\omega$, the same is not clear for the complete reconstruction using an orthonormal basis of $L^2(\Omega)$. Indeed, one cannot even expect in general that evaluating the right hand side of (8) for an arbitrarily perturbed $\widetilde{y}\chi_\omega$ and the elements of an orthonormal basis φ_k defines a square-summable series. We therefore investigate the question of stability in section 5 numerically.

1.3 Outline of numerical forward assimilation

Using equation (8), we can calculate projections of the final state $y(T)$ from the given data and measurements only. This suggests the following approach for numerically computing an approximation $y^H(T)$ of $y(T)$: First, we fix a finite dimensional subspace $V_H \subset \mathrm{L}^2(\Omega)$. To calculate the projection $y^H(T)$ of $y(T)$ into V_H (with respect to the $\mathrm{L}^2(\Omega)$ inner product), we take an orthonormal basis $\{\Phi_k\}_{k=1}^n$, compute the coefficients $\langle y(T), \Phi_k \rangle_{\mathrm{L}^2(\Omega)}$, and form the sum

$$y^H(T) := \sum_{k=1}^{n} \langle y(T), \Phi_k \rangle_{\mathrm{L}^2(\Omega)} \, \Phi_k. \tag{9}$$

We approximate the projections onto each Φ_k by numerically solving a penalized formulation of the null-control problem for (2) with $\varphi_T = \Phi_k$. The penalty parameter $\beta > 0$ can be chosen arbitrary and is independent from the measurement data $y\chi_\omega$. It is possible to show convergence of the corresponding optimal controls $v^\beta(\Phi_k)$ to the exact control $v(\Phi_k)$ given by Corollary 1.4 when $\beta \to 0$. As will be detailed in section 2, the solution of the control problem can be found by solving an operator equation involving a pair of partial differential equations. The numerical solution of this operator equation is done by an iterative method for a Galerkin approximation in a finite dimensional space V_h, which needs to embed V_H, and a discontinuous Galerkin method in time. (Using standard finite element spaces for V_H and V_h, for example, the embedding is ensured by the condition $h < H$ on iteratively refined grids.) In section 3 we will show existence and uniqueness of discrete controls $v^{h,\beta}(\Phi_k)$ converging to $v^\beta(\Phi_k)$. If f and $y\chi_\omega$ are projected into the finite dimensional space as well, the inner products in (9) can then be calculated exactly. Furthermore, we can show convergence for the numerical approximations of $\langle y(T), \Phi_k \rangle_{\mathrm{L}^2(\Omega)}$ to the exact coefficients for $h, \beta \to 0$.

The convergence of the approximations $y^H(T)$ of $y(T)$ as $H \to 0$ is treated in section 4.1. It remains to specify the orthonormal basis $\{\Phi_k\}_{k=1}^n$ of V_H. The key of our approach is to make use of a proper orthogonal decomposition for this purpose. For an efficient assimilation process, it is essential to employ these basis elements in order of their importance for the approximation of $y(T)$. This is achieved by first estimating the coefficients $\langle y^H(T), \Phi_k \rangle_{\mathrm{L}^2(\Omega)}$ via an interpolant of the measurements, which also provides a (heuristic) error estimate and hence an adaptive termination criterion for the summation. The remainder of section 4 is devoted to this issue.

Finally, we give in section 5 details on the implementation and present numerical examples in two and three dimensions which demonstrate the effectiveness of the proposed method.

2 Control of the adjoint equation

The solution of the forward data assimilation requires computing null-controls for the adjoint equation. To motivate the numerical method, we recall here the relevant basic

facts from exact controllability, including its approximation via penalization. We refer to, e.g., [11] for a complete exposition. Since we know a priori that the desired null control exists, we can additionally show convergence of the penalized approximations.

As will be seen below (cf. Remark 2.8), it is necessary to use $\varphi(T) = 0$ as the final value of the adjoint equation (2). Since the equation is linear, this can be easily achieved by solving the partial differential equation for given final value φ_T with the control $v = 0$. The solution $\widetilde{\varphi}$ then defines a new target state $\varphi_0 := -\widetilde{\varphi}(0)$ at time $t = 0$, which is reachable from $\varphi(T) = 0$ by the same control that drives φ_T to zero.

We now give a precise formulation of the control problem for the adjoint equation:

Problem 2.1. *Given $\omega \subset \Omega$ and $\varphi_0 \in \mathrm{L}^2(\Omega)$, find*

$$\inf_{v \in \mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))} \frac{1}{2} \|v\|^2_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \quad \text{s.t. } \varphi(0;v) = \varphi_0,$$

where $\varphi_T = 0$ and $\varphi(0;v)$ denotes the solution of (2) corresponding to the control v, evaluated at $t = 0$.

We are thus specifically searching for the control of minimal $\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))$ norm. Due to estimate (5) in Proposition 1.1, the final value $\varphi(0;v)$ is continuous in v. The set of all controls v satisfying $\varphi(0;v) = \varphi_0$ is therefore closed and convex. It is also non-empty, since by Theorem 1.4 there exists at least one such control. Thus, using standard results from functional analysis, Problem 2.1 has a unique solution, which we denote with $\overline{v}$. The stability of the control problem follows immediately from estimate (7).

2.1 Penalized approximation

Although there is a unique solution of the exact control problem we are considering, this is no longer clear for discretized versions of Problem 2.1. It is therefore numerically convenient to consider a penalized version of the control problem:

Problem 2.2. *Given $\omega \subset \Omega$, $\varphi_0 \in \mathrm{L}^2(\Omega)$ and $\beta > 0$, find*

$$\inf_{v \in \mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))} \left\{ \frac{1}{2} \|v\|^2_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} + \frac{1}{2\beta} \|\varphi(0;v) - \varphi_0\|^2_{\mathrm{L}^2(\Omega)} \right\},$$

where $\varphi_T = 0$ and $\varphi(0;v)$ denotes again the solution of (2) with control v at $t = 0$.

By the same arguments as above, this second control problem also has a unique solution, which we denote with $\overline{v}^\beta$. The following theorem shows that we have convergence of $\overline{v}^\beta$ to the exact control $\overline{v}$ for $\beta \to 0$. This implies that we are allowed to choose $\beta > 0$ arbitrarily small, independent from the assimilation problem itself. In particular, the choice of β will be independent of the measurement data h.

Theorem 2.3. *The solutions $\overline{v}^\beta$ of the penalized control problem 2.2 converge to the solution $\overline{v}$ of the exact control problem 2.1:*

$$\lim_{\beta\to 0} \overline{v}^\beta = \overline{v}.$$

Proof. We first note that by definition of the control problems we have

$$\frac{1}{2}\left\|\overline{v}^\beta\right\|^2_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} + \frac{1}{2\beta}\left\|\varphi(0;\overline{v}^\beta) - \varphi_0\right\|^2_{\mathrm{L}^2(\Omega)} \leq \frac{1}{2}\left\|\overline{v}\right\|^2_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}$$

Multiplying by $\beta > 0$, this directly yields

$$\lim_{\beta\to 0} \varphi(0;\overline{v}^\beta) = \varphi_0.$$

Moreover, the controls $\overline{v}^\beta\chi_\omega$ are bounded by $\overline{v}\chi_\omega$:

$$\left\|\overline{v}^\beta\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \leq \left\|\overline{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}. \tag{10}$$

Thus, we can find a weakly converging subsequence

$$\overline{v}^{\beta_k} \rightharpoonup \widetilde{v},$$

where $\lim_{k\to\infty}\beta_k = 0$. Since $v \mapsto \varphi(0;v)$ is a continuous linear mapping, we have

$$\varphi(0;\widetilde{v}) = \varphi_0,$$

i.e. $\widetilde{v}$ is an exact control. From (10) and the weak semi-continuity of the norm, we obtain

$$\left\|\widetilde{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \leq \liminf_{k\to\infty}\left\|\overline{v}^{\beta_k}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \leq \left\|\overline{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}.$$

On the other hand, using the optimality of $\overline{v}$ yields $\left\|\overline{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \leq \left\|\widetilde{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}$. Therefore, $\widetilde{v}$ is an exact control of minimal norm. Due to uniqueness for the exact control problem 2.1, $\widetilde{v} = \overline{v}$ holds.

The next step is to show strong convergence for the sequence v^{β_k}. To this end we use again the boundedness (10) of the controls. We can assume (after possibly extracting a subsequence) that

$$\lim_{k\to\infty}\left\|\overline{v}^{\beta_k}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} = K \leq \left\|\overline{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}.$$

Since the closed convex set $B_K(0)$ (the ball of radius K in $\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))$) is also weakly closed, we can pass to the limit and obtain

$$\left\|\widetilde{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \leq K \leq \left\|\overline{v}\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}.$$

Together with $\widetilde{v} = \overline{v}$ this yields $K = \|\overline{v}\|_{L^2(0,T;L^2(\omega))}$. The sequence $\overline{v}^{\beta_k}$ is therefore weakly convergent to $\overline{v}$ and the norms $\|\overline{v}^{\beta_k}\|_{L^2(0,T;L^2(\omega))}$ converge to $\|\overline{v}\|_{L^2(0,T;L^2(\omega))}$. As $L^2(0,T;L^2(\omega))$ is a Hilbert space, this implies strong convergence.

It remains to show convergence of any other subsequence $\overline{v}^{\beta_l}$ with $\lim_{l\to\infty} \beta_l = 0$. Suppose there is a sequence $\overline{v}^{\beta_l}$ which is *not* converging to $\overline{v}$. Without loss of generality (again by extracting a subsequence), there is an $\varepsilon > 0$ such that

$$\left\|\overline{v}^{\beta_l} - \overline{v}\right\|_{L^2(0,T;L^2(\omega))} > \varepsilon, \qquad \text{for all } l = 1, 2, \ldots.$$

On the other hand, we can employ the same arguments as above to find a subsequence of $(\overline{v}^{\beta_l})_l$ strongly converging to $\overline{v}$, which gives a contradiction. □

2.2 Characterization of control

The solution of Problem 2.2 can be characterized using extremality relations in Fenchel duality, which yields an efficient numerical method. Hence, we introduce the adjoint of the equation to be controlled, which is itself the adjoint of the state equation. To reduce the risk of confusion, we will use the term *biadjoint* to describe this equation and the corresponding solution:

$$\begin{cases} \partial_t \psi - \nabla \cdot (c \nabla \psi) + b^T \nabla \psi = 0, & \text{in } \Omega \times (0,T), \\ \psi = 0, & \text{on } \Gamma \times [0,T], \\ \psi(\cdot, 0) = \psi_0, & \text{in } \Omega. \end{cases} \tag{11}$$

Like the adjoint equation (2), this partial differential equation has a unique solution $\psi \in L^2(0,T;H_0^1(\Omega))$. The regularity of ψ depends on the smoothness of ψ_0:

Proposition 2.4. *For given $\psi_0 \in L^2(\Omega)$, equation (11) has a unique weak solution, which satisfies*

$$\psi \in L^2(0,T;H_0^1(\Omega)) \cap C(0,T;L^2(\Omega)), \qquad \partial_t \psi \in L^2(0,T;H^{-1}(\Omega)).$$

If in addition $\psi_0 \in H^1(\Omega)$ holds, this solution satisfies

$$\begin{cases} \psi \in L^2(0,T;H^2(\Omega)) \cap C(0,T;H^1(\Omega)) \text{ and} \\ \partial_t \psi \in L^2(0,T;H^1(\Omega)) \end{cases} \tag{12}$$

Proof. The statements follow from the same regularity results for parabolic problems given in the proof of Proposition 1.1. □

The connection between the adjoint and biadjoint equation can be expressed in operator notation as follows. We first define the operator

$$\Lambda : \begin{cases} L^2(\Omega) \to L^2(\Omega), \\ \psi_0 \mapsto \varphi(0; \psi \chi_\omega). \end{cases} \tag{13}$$

That is, for a given function ψ_0, we solve the biadjoint equation (11), use this solution as the control $v = \psi\chi_\omega$ of the adjoint equation (2), and evaluate its solution φ at $t = 0$.

The following isometry property for the operator Λ is essential for all further arguments.

Lemma 2.5. *For every $\psi_0, \widetilde{\psi}_0 \in L^2(\Omega)$ we have*

$$\langle \Lambda\psi_0, \widetilde{\psi}_0 \rangle_{L^2(\Omega)} = \langle \psi, \widetilde{\psi} \rangle_{L^2(0,T;L^2(\omega))},$$

where ψ and $\widetilde{\psi}$ are the solutions to the biadjoint equation (11) with initial value ψ_0 and $\widetilde{\psi}_0$ respectively.

Proof. Let φ be the solution of the adjoint equation (2) corresponding to the control $\psi\chi_\omega$. Multiplying the biadjoint equation (11) with the initial value $\widetilde{\psi}_0$ with φ and applying Green's formula yields

$$\begin{aligned} 0 &= \int_0^T \int_\Omega (\partial_t \widetilde{\psi} - \nabla \cdot (c\nabla\widetilde{\psi}) + b^T \nabla \widetilde{\psi})\varphi \, dxdt \\ &= \int_0^T \int_\Omega \widetilde{\psi}(-\partial_t\varphi - \nabla \cdot (c\nabla\varphi) - \nabla \cdot (b\varphi)) \, dxdt - \int_\Omega \widetilde{\psi}_0 \varphi(0) dx, \end{aligned}$$

where we have used zero boundary conditions on φ. The claim now follows from the adjoint equation and the fact that by construction $\varphi(0) = \Lambda\psi_0$. □

Proposition 2.6. *The operator $\Lambda : L^2(\Omega) \to L^2(\Omega)$ defined by (13) is a continuous linear operator which is self-adjoint and positive semi-definite.*

Proof. Linearity follows directly from the definition of Λ and linearity of the partial differential equations (11) and (2). Due to Lemma 2.5, we have

$$\langle \Lambda\psi_0, \widehat{\psi}_0 \rangle_{L^2(\Omega)} = \int_0^T \int_\omega \psi\widehat{\psi} \, dxdt = \langle \psi_0, \Lambda\widehat{\psi}_0 \rangle_{L^2(\Omega)} \tag{14}$$

for every $\psi_0, \widehat{\psi}_0 \in L^2(\Omega)$. According to the Hellinger-Töplitz theorem, Λ is therefore continuous and self-adjoint. Positive semi-definiteness is also a direct consequence of equation (14). □

It is now possible to characterize the solution of Problem 2.2:

Proposition 2.7. *The operator equation*

$$(\Lambda + \beta I)\psi_0 = \varphi_0 \tag{15}$$

has a unique solution $\overline{\psi}_0^\beta \in L^2(\Omega)$. The unique solution of Problem 2.2 is given by $v^\beta = \overline{\psi}^\beta\chi_\omega$, where $\overline{\psi}^\beta$ is the solution of (11) with initial value $\overline{\psi}_0^\beta$.

Proof. We only sketch the idea of the proof here and refer the reader to [11] for a more detailed discussion. The main technique used is the duality theory by Fenchel and Rockafellar [21], which can be used to show

$$\min_{v\in L^2(0,T;L^2(\omega))} \left\{ \frac{1}{2} \|v\|^2_{L^2(0,T;L^2(\omega))} + \frac{1}{2\beta} \|\varphi(0;v) - \varphi_0\|^2_{L^2(\Omega)} \right\} \tag{16}$$
$$= \max_{\psi_0\in L^2(\Omega)} \left\{ \langle \psi_0, \varphi_0\rangle_{L^2(\Omega)} - \frac{1}{2}\beta \|\psi_0\|^2_{L^2(\Omega)} - \frac{1}{2} \|\psi\|^2_{L^2(0,T;L^2(\omega))} \right\}$$

for every $\beta > 0$. The same theory gives us the existence of unique, finite solutions $\overline{v}^\beta$ and $\overline{\psi}^\beta_0$ for these problems. Using the equality of the extrema and Lemma 2.5, we get

$$\frac{1}{2} \left\| \overline{v}^\beta - \overline{\psi}^\beta \right\|^2_{L^2(0,T;L^2(\omega))} =$$
$$\left\langle \overline{\psi}^\beta_0, \varphi_0 - \varphi(0;\overline{v}^\beta) \right\rangle_{L^2(\Omega)} - \frac{1}{2}\beta \left\| \overline{\psi}^\beta_0 \right\|_{L^2(\Omega)} - \frac{1}{2\beta} \left\| \varphi_0 - \varphi(0;\overline{v}^\beta) \right\|^2_{L^2(\Omega)} \leq 0.$$

The upper bound follows from the Cauchy-Schwarz inequality together with Young's inequality. We thus have

$$\overline{v}^\beta = \overline{\psi}^\beta \chi_\omega \qquad \text{for all } \beta > 0.$$

It remains to show that $\overline{\psi}^\beta_0$ is the unique solution of the operator equation (15). Since $\overline{\psi}^\beta_0$ solves the optimization problem on the right hand side of (16), Lemma 2.5 yields

$$\overline{\psi}^\beta_0 = \operatorname{argmin}_{\psi_0\in L^2(\Omega)} \left\{ \langle (\Lambda + \beta I)\psi_0, \psi_0\rangle_{L^2(\Omega)} - \langle \psi_0, \varphi_0\rangle_{L^2(\Omega)} \right\}.$$

The operator Λ is positive semi-definite by Proposition 2.6. Therefore, $(\Lambda + \beta I)$ is strictly positive definite and the minimization problem above is equivalent to

$$\left\langle (\Lambda + \beta I)\overline{\psi}^\beta_0, w \right\rangle_{L^2(\Omega)} = \langle \varphi_0, w\rangle_{L^2(\Omega)} \qquad \text{for all } w \in L^2(\Omega).$$

This yields (15). Uniqueness of the solution is then a direct consequence of the positive definiteness. □

Remark 2.8. *For the original formulation of the null controllability problem as given in Corollary 1.4, the operator Λ defined by (13)—and hence $\Lambda + \beta I$—would be affine, not linear. This would make the proof of Proposition 2.7, as well as the numerical solution of (15), more involved.*

Since $\Lambda + \beta I$ is linear, self-adjoint and positive definite, an efficient approach for the numerical solution of equation (15) is to apply the method of conjugated gradients (CG) to a suitable discretization of this operator. This will be justified and discussed in the next section.

3 Numerical calculation of projection

In this section, we detail a numerical method for the approximation of the projection coefficient $\langle y(T), \varphi_T \rangle_{L^2(\Omega)}$ in formula (8). The main task here is the numerical solution of the control problem 2.2 presented in the previous section.

Due to Proposition 2.7, the control problem reduces to solving the operator equation (15). Since the operator Λ involves solutions of the adjoint and biadjoint equation, we first introduce an appropriate discretization scheme for these partial differential equations in section 3.1. In particular, the spatial discretization leads to finite dimensional subspaces $V_h \subset H_0^1(\Omega)$. This allows us to formulate a discrete version Λ^h of operator Λ, defined on V_h. We are then able to show the solvability of the discrete analog of equation (15) and detail an algorithm for this task in section 3.2.

Finally, we prove convergence of the corresponding discrete approximations of the coefficients $\langle y(T), \varphi_T \rangle_{L^2(\Omega)}$ for given $\varphi_T \in V_H$, where $H > h$ is fixed, to the exact coefficient when $h, \beta \to 0$.

3.1 Discretization and operator convergence

We use a variational discretization consisting of standard conforming finite elements in space and discontinuous piecewise polynomials in time, which corresponds to the classical discontinuous Galerkin (DG) method for parabolic problems (cf. [26] and references therein). This choice satisfies a discrete isometry property which is crucial for the solvability of the discrete problem, while having the necessary approximation properties to ensure convergence of the discrete solution to the sought control. Details of the implementation are given in section 5.1.

For the spatial discretization, we take as ansatz and trial space

$$V_h = \operatorname{span}\{B_1, \dots, B_n\} \subset H_0^1(\Omega),$$

the space generated by the finite element form functions B_j which are globally continuous piecewise polynomials of degree at most s on a quasi-uniform mesh of size h. We denote by $\mathcal{P}_h : L^2(\Omega) \to V_h$ the orthogonal projection from $L^2(\Omega)$ onto V_h.

Defining a partition $0 = t_0 < t_1 < \cdots < t_N = T$ of $[0, T]$, we introduce the space

$$S_r := \{v \in L^2(0, T; H^1(\Omega));\ v\chi_{I_m} \in \Pi^r(t_{m-1}, t_m; V_h)\ m = 1, \dots, N\},$$

where $I_m := (t_{m-1}, t_m]$ and $\Pi^r(t_{m-1}, t_m; V_h)$ is the space of polynomials of degree at most r having values in V_h. Similarly, we define

$$S_r^* := \{v \in L^2(0, T; H^1(\Omega));\ v\chi_{J_m} \in \Pi^r(t_{m-1}, t_m; V_h),\ m = 1, \dots, N\}$$

with $J_m := [t_{m-1}, t_m)$.

Setting $v_m^+ := \lim_{t \to t_m^+}$, $v_m^- := \lim_{t \to t_m^-}$, $[v]_m := v_m^+ - v_m^-$ (note that these definitions

are independent of whether v is in S_r^* or S_r), and

$$\begin{cases} a : [0,T] \times \mathrm{H}_0^1(\Omega) \times \mathrm{H}_0^1(\Omega) \to \mathbb{R}, \\ a(t;u,v) := \langle c\nabla u, \nabla v\rangle_{\mathrm{L}^2(\Omega)} + \langle b\nabla u, v\rangle_{\mathrm{L}^2(\Omega)}, \end{cases}$$

$$\begin{cases} a^* : [0,T] \times \mathrm{H}_0^1(\Omega) \times \mathrm{H}_0^1(\Omega) \to \mathbb{R}, \\ a^*(t;u,v) := \left\langle c^T\nabla u, \nabla v\right\rangle_{\mathrm{L}^2(\Omega)} + \langle bu, \nabla v\rangle_{\mathrm{L}^2(\Omega)}, \end{cases}$$

we introduce the discrete variational form of problem (11): Find $\psi \in S_r$ such that

$$\begin{cases} \sum_{m=1}^{N} \left[\int_{I_m} \Big(\langle \partial_t \psi, v\rangle_{\mathrm{L}^2(\Omega)} + a(t;\psi,v) \Big)\, dt + \big\langle [\psi]_{m-1}, v_{m-1}^+ \big\rangle_{\mathrm{L}^2(\Omega)} \right] = 0, \\ \qquad\qquad \psi_0^- = \psi_0^h \end{cases} \tag{17}$$

holds for all $v \in S_r$.

The discrete variational form of the adjoint problem (2) with control $\psi\chi_\omega$ is defined as: Find $\varphi \in S_r^*$ such that

$$\begin{cases} \sum_{m=1}^{N} \left[\int_{J_m} \Big(\langle -\partial_t \varphi, u\rangle_{\mathrm{L}^2(\Omega)} + a^*(t;\varphi,u) - \langle \psi, u\rangle_{\mathrm{L}^2(\omega)} \Big)\, dt - \big\langle [\varphi]_m, u_m^- \big\rangle_{\mathrm{L}^2(\Omega)} \right] = 0, \\ \qquad\qquad \varphi_N^+ = 0 \end{cases} \tag{18}$$

holds for all $u \in S_r^*$. Note that with the substitution $\tau_m = t_N - t_m$ (i.e., integrating backwards in time), we recover the setting of (17).

These discretizations can be reformulated as time stepping schemes, which for $r = 0$ are equivalent to the implicit Euler method (if the time integration is approximated suitably).

Remark 3.1. *For highly convection-dominated problems, the equations should be considered in Lagrangian variables. Since the focus of this work is on the feasibility of the proposed data assimilation approach, we restrict ourselves to problems which are moderately convection-dominated. For such problems, using a sufficiently fine grid as well as enforcing the condition $h/2 \leq \Delta t \leq h$ is sufficient to guarantee stability, at the cost of increased computational work. Similarly, we do not address the issue of adaptive refinement in space and time, although this is possible using the discretization given above.*

For the remainder of the section, we always assume that Δt is sufficiently small compared to h to ensure stability of the finite element method. This also means that we can treat $\{h, \Delta t\}$ as a single sequence in the following, indicated by the superscript h.

The full discretization can now be used to compute approximations to $\Lambda\psi_0^h$ for any $\psi_0^h \in V_h$. This is equivalent to applying a discrete version

$$\Lambda^h : V_h \to V_h$$

of the operator Λ to the value ψ_0^h. This discretized operator is still linear and (since it is finite dimensional) continuous. In the following, we denote by ψ^h the solution of problem (17) corresponding to the initial value ψ_0^h. Similarly, $\varphi_m^h(\psi^h\chi_\omega)$ stands for the solution of problem (18) corresponding to the control $\psi^h\chi_\omega$ at time step t_m.

We close this section with the convergence of the discontinuous Galerkin discretization of the operator Λ.

Proposition 3.2. *For $\psi_0 \in \mathrm{H}_0^1(\Omega)$ and $\psi_0^h := \mathcal{P}_h\psi_0$, the following holds:*

$$\lim_{h\to 0} \sup_{0\le t\le T} \|\psi(t) - \psi^h(t)\|_{\mathrm{L}^2(\Omega)} = 0 \tag{19}$$

and

$$\lim_{h\to 0} \|\Lambda\psi_0 - \Lambda^h\psi_0^h\|_{\mathrm{L}^2(\Omega)} = 0. \tag{20}$$

Proof. Proof of (19). By [7, Th. 3.2], we can bound the approximation error by the projection error

$$\sup_{0\le t\le T} \|\psi - \psi^h\|_{\mathrm{L}^2(\Omega)} \le C(T)\, \|\psi - \mathbb{P}_r\psi\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))},$$

since $\psi_0^h = \mathcal{P}_r\psi_0$ and the same subspace V_h is used in every time step. Here, $\mathbb{P}_r$ is the local projection onto S_r defined by

$$(\mathbb{P}_r u)(t_m) = \mathcal{P}_h(u(t_m)), \qquad \int_{t_{m-1}}^{t_m} \langle u - \mathbb{P}_r u, v\rangle_{\mathrm{L}^2(\Omega)} = 0$$

for all m and all $v \in \Pi^{r-1}(t_{m-1}, t_m; V_h)$, which is exact for all $u \in \Pi^r(t_{m-1}, t_m; V_h)$ (cf. [26, Th. 12.1]). From standard approximation properties, we therefore obtain for $s \ge 1$ and $r \ge 0$:

$$\sup_{0\le t\le T} \|\psi(t) - \psi^h(t)\|_{\mathrm{L}^2(\Omega)} \le C\left(h\,\|\psi\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))} + \sqrt{\Delta t}\,\|\partial_t\psi\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))}\right).$$

Since the norms on the right hand side are bounded by the regularity estimate (12), the convergence as $\{h, \Delta t\} \to 0$ follows.

Proof of (20). By definition, $\Lambda\psi_0 = \varphi(0;\psi\chi_\omega)$, and correspondingly, $\Lambda^h\psi_0^h = \varphi_0^+(\psi^h\chi_\omega)$. We split the approximation error as follows:

$$\|\Lambda\psi_0 - \Lambda^h\psi_0^h\|_{\mathrm{L}^2(\Omega)} \le \left\|\varphi(0;\psi\chi_\omega) - \varphi(0;\psi^h\chi_\omega)\right\|_{\mathrm{L}^2(\Omega)} + \left\|\varphi(0;\psi^h\chi_\omega) - \varphi_0^+(\psi^h\chi_\omega)\right\|_{\mathrm{L}^2(\Omega)}$$

To show convergence of the first term, we use the linearity of (2) in the right hand side and estimate (5) to obtain the bound

$$\begin{aligned}\left\|\varphi(0;\psi\chi_\omega) - \varphi(0;\psi^h\chi_\omega)\right\|_{\mathrm{L}^2(\Omega)} &= \left\|\varphi(0;(\psi-\psi^h)\chi_\omega)\right\|_{\mathrm{L}^2(\Omega)} \\ &\le C\left\|\psi - \psi^h\right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \le C(T)\sup_{0\le t\le T}\|\psi(t) - \psi^h(t)\|_{\mathrm{L}^2(\omega)} \to 0\end{aligned} \tag{21}$$

according to (19).

The convergence of the second term follows (after time reversal) from the same arguments as in the proof of (19), making use of the corresponding a priori estimates for (11):

$$\|\varphi(0;\psi^h\chi_\omega) - \varphi_0^+(\psi^h\chi_\omega)\|_{L^2(\Omega)} \le C\left(h\,\|\varphi\|_{L^2(0,T;H^1(\Omega))} + \sqrt{\Delta t}\,\|\partial_t\varphi\|_{L^2(0,T;H^1(\Omega))}\right) \tag{22}$$

□

Remark 3.3. *Similar error estimates hold for the approximation error, if different subspaces V_h^m are used in different time steps (e.g., for adaptive refinement), cf. [7].*

3.2 Solvability of discrete control problem

Consider the discrete analog of equation (15),

$$\Lambda^{h,\beta}\psi_0^h := (\Lambda^h + \beta I)\psi_0^h = \varphi_0^h, \tag{23}$$

where φ_0^h is a numerical approximation of the shifted target φ_0 (as introduced at the beginning of section 2) in the finite element space V_h.

We will show now that (23) has a unique solution $\overline{\psi}_0^{h,\beta}$ for every fixed space-time grid $\{h, \Delta t\}$. This follows directly from the following isometry property:

Lemma 3.4. *For every $\psi_0^h, \widetilde{\psi}_0^h \in V_h$, we have*

$$\left\langle \Lambda^h\psi_0^h, \widetilde{\psi}_0^h \right\rangle_{L^2(\Omega)} = \left\langle \psi^h, \widetilde{\psi}^h \right\rangle_{L^2(0,T;L^2(\omega))}, \tag{24}$$

hence Λ^h is symmetric and positive semi-definite.

Proof. In terms of the discretized adjoint and biadjoint equation, we have to show that

$$\langle \varphi_0^+, \widetilde{\psi}_0^- \rangle_{L^2(\Omega)} = \langle \psi, \widetilde{\psi} \rangle_{L^2(0,T;L^2(\omega))},$$

holds for the discrete solutions $\psi, \widetilde{\psi} \in S_r$ and $\varphi = \varphi(t, \psi\chi_\omega) \in S_r^*$ (for convenience, we drop the superscript h for the duration of this proof).

We start with the following identity, which holds for all $u, v \in S_r \cup S_r^*$:

$$\begin{aligned} s_m &:= \int_{I_m} \langle \partial_t u, v\rangle_{L^2(\Omega)} + a(t;u,v)\,dt + \left\langle [u]_{m-1}, v_{m-1}^+ \right\rangle_{L^2(\Omega)} \\ &= \int_{J_m} \langle -\partial_t v, u\rangle_{L^2(\Omega)} + a^*(t;v,u)\,dt + \left\langle u_m^-, v_m^- \right\rangle_{L^2(\Omega)} - \left\langle u_{m-1}^+, v_{m-1}^+ \right\rangle_{L^2(\Omega)} \\ &\qquad + \left\langle u_{m-1}^+, v_{m-1}^+ \right\rangle_{L^2(\Omega)} - \left\langle u_{m-1}^-, v_{m-1}^+ \right\rangle_{L^2(\Omega)} \\ &= \int_{J_m} \langle -\partial_t v, u\rangle_{L^2(\Omega)} + a^*(t;v,u)\,dt + \left\langle u_m^-, v_m^- \right\rangle_{L^2(\Omega)} - \left\langle u_{m-1}^-, v_{m-1}^+ \right\rangle_{L^2(\Omega)} \end{aligned} \tag{25}$$

by partial integration, and since $\int_{I_m} f(t)dt = \int_{t_{m-1}}^{t_m} f(t)dt = \int_{J_m} f(t)dt$ holds for all f by elementary properties of the Lebesgue integral.

Now we observe that we are allowed to test equation (17) with $\varphi \in S_r^*$ and equation (18) with $\psi \in S_r$, since $\hat{\varphi}(t) := \varphi(t)$ for $t \in (t_{m-1}, t_m)$, $\hat{\varphi}(t_m) := \varphi_m^-$ satisfies $\hat{\varphi} \in S_r$,

$$\int_{I_n} \langle \partial_t \psi, \varphi \rangle_{\mathrm{L}^2(\Omega)} + a(t; \psi, \varphi)\, dt = \int_{I_n} \langle \partial_t \psi, \hat{\varphi} \rangle_{\mathrm{L}^2(\Omega)} + a(t; \psi, \hat{\varphi})\, dt,$$

and $\varphi_{m-1}^+ = \hat{\varphi}_{m-1}^+$ for every $\varphi \in S_r^*$ (and similarly, we can construct a $\hat{\psi}$ for ψ).

We may thus substitute $u = \widetilde{\psi} \in S_r$ and $v = \varphi \in S_r^*$ in (25), where $\widetilde{\psi}$ is the solution of (17) corresponding to the initial value $\widetilde{\psi}_0$ and φ is the solution of (18) corresponding to the control $\psi\chi_\omega$. Summing over $m = 1, \dots, N$ and rearranging the jump terms yields:

$$\begin{aligned}
0 &= \sum_{m=1}^{N} s_m - \langle \widetilde{\psi}_N^-, \varphi_N^+ \rangle_{\mathrm{L}^2(\Omega)} \\
&= \sum_{m=1}^{N} \int_{J_m} \langle -\partial_t \varphi, \widetilde{\psi} \rangle_{\mathrm{L}^2(\Omega)} + a^*(t; \varphi, \psi)\, dt + \langle \widetilde{\psi}_m^-, \varphi_m^- \rangle_{\mathrm{L}^2(\Omega)} - \langle \widetilde{\psi}_m^-, \varphi_m^+ \rangle_{\mathrm{L}^2(\Omega)} \\
&\quad - \langle \widetilde{\psi}_0^-, \varphi_0^+ \rangle_{\mathrm{L}^2(\Omega)} \\
&= \sum_{m=1}^{N} \int_{J_m} \langle -\partial_t \varphi, \widetilde{\psi} \rangle_{\mathrm{L}^2(\Omega)} + a^*(t; \varphi, \widetilde{\psi})\, dt - \langle \widetilde{\psi}_m^-, [\varphi]_m \rangle_{\mathrm{L}^2(\Omega)} - \langle \widetilde{\psi}_0^-, \varphi_0^+ \rangle_{\mathrm{L}^2(\Omega)} \\
&= \sum_{n=1}^{N} \int_{J_m} \langle \psi, \widetilde{\psi} \rangle_{\mathrm{L}^2(\omega)}\, dt - \langle \varphi_0^+, \widetilde{\psi}_0^- \rangle_{\mathrm{L}^2(\Omega)}.
\end{aligned}$$

Symmetry of Λ^h follows immediately from equation (24) by switching $\widetilde{\psi}_0^h$ and ψ_0^h, and positive semi-definiteness by setting $\widetilde{\psi}_0^h = \psi_0^h$:

$$\left\langle \Lambda^h \psi_0^h, \psi_0^h \right\rangle_{\mathrm{L}^2(\Omega)} = \left\| \psi^h \right\|^2_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} \geq 0.$$

□

Hence, $\Lambda^{h,\beta}$ is a linear, symmetric, positive definite, finite dimensional operator, and therefore the discrete operator equation (23) has a unique solution, which we denote by $\overline{\psi}_0^{h,\beta}$. As noted above, due to the positive definiteness of $\Lambda^{h,\beta}$, the CG method for the iterative computation of this solution is a natural choice. Algorithm 1 details the necessary steps.

Remark 3.5. *We choose the initial value $\psi_0^{(1)} = 0$ if no better guess is available. The stopping criterion for the CG method can be a combination of absolute and relative tolerance with respect to the target value φ_0^h. Note that while $\Lambda^{h,\beta}$ is symmetric positive definite, we can only guarantee semi-definiteness for Λ^h. For small β, the CG algorithm above should be replaced with other Krylov methods such as BiCGstab, which in our experiments gave better performance.*

Algorithm 1 Computation of control

compute solution $\widetilde{\varphi}$ of adjoint equation with initial value φ_T and control $v = 0$
shifted target: $\varphi_0^h := -\widetilde{\varphi}(0)$
choose initial value $\psi_0^{(1)} \in V_h$
compute solution $\psi^{(1)}$ of biadjoint equation with initial value $\psi_0^{(1)}$
compute solution $\varphi^{(1)}$ of adjoint equation with control $\psi^{(1)}\chi_\omega$
operator value: $\varphi_0^{(1)} := \Lambda^h \psi_0^{(1)}$
residual: $g^{(1)} := (\varphi_0^{(1)} + \beta\psi_0^{(1)}) - \varphi_0^h$
conjugate direction: $w^{(1)} := g^{(1)}$
for $k = 1, 2, \ldots$ **do** {perform CG method}
  operator value: $\varphi_0^{(k+1)} = \Lambda^h w^{(k)}$
  step length: $\rho^{(k)} := \|g^{(k)}\|^2_{\mathrm{L}^2(\Omega)} / \langle \varphi_0^{(k+1)} + \beta\psi_0^{(k)}, w^{(k)} \rangle_{\mathrm{L}^2(\Omega)}$
  residual: $g^{(k+1)} := g^{(k)} - \rho^{(k)}(\varphi_0^{(k+1)} + \beta\psi_0^{(k)})$
  new vector: $\psi^{(k+1)} := \psi^{(k)} - \rho^{(k)} w^{(k)}$
  if $\|g^{(k+1)}\|_{\mathrm{L}^2(\Omega)} \leq tol \cdot \|\varphi_0^h\|_{\mathrm{L}^2(\Omega)}$ **then**
    break;
  end if
  conjugate direction: $w^{(k+1)} := g^{(k+1)} + (\|g^{(k+1)}\|^2_{\mathrm{L}^2(\Omega)} / \|g^{(k)}\|^2_{\mathrm{L}^2(\Omega)}) w^{(k)}$
end for
compute solution $\psi^{(k)}$ of biadjoint equation with initial value $\psi_0^{(k)}$
return $\psi^{(k)}\chi_\omega$

3.3 Convergence in $h \to 0$

We will now show convergence of the discrete control, computed with Algorithm 1, to the exact control given by Corollary 1.4. To this end, we fix $H > 0$ and an arbitrary vector φ_T from the corresponding finite dimensional subspace $V_H \subset \mathrm{H}_0^1(\Omega)$. As before, the vector φ_0 denotes the shifted target value corresponding to our initial value φ_T and φ_0^h its numerical approximation in $V_h \supset V_H$. Specifically, it is the negative of the numerical solution of the adjoint equation with initial value φ_T and control $v = 0$, evaluated at time $t = 0$. Since $\varphi_T \in V_h$, by a similar argument as in Proposition 3.2 (cf. especially estimate (22)), we have that

$$\lim_{h \to 0} \varphi_0^h = \varphi_0.$$

In the following, we assume that $V_{h_2} \subset V_{h_1}$ for any $h_1 \leq h_2 \leq H$. This can be achieved by using standard finite elements on successively refined grids, where h denotes the mesh size of the largest element. Thus, the application of $\Lambda^{h,\beta}$ to an element of V_H is well-defined for any $h < H$.

The following theorem shows the convergence of the discrete control to the solution of Problem 2.2 as $h \to 0$.

Theorem 3.6. *Let $\overline{\psi}_0^{h,\beta}$ be the unique solution of (23) and $\overline{\psi}_0^{\beta}$ the solution of (15). Then the following holds for the solutions $\overline{\psi}^{h,\beta}$ of (17) and $\overline{\psi}^{\beta}$ of (11) with initial value $\overline{\psi}_0^{h,\beta}$ and $\overline{\psi}_0^{\beta}$, respectively:*

$$\lim_{h\to 0} \left\| \overline{\psi}^{h,\beta} - \overline{\psi}^{\beta} \right\|_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))} = 0.$$

Proof. The proof is identical to [11, Th. 1.3], using Proposition 3.2 in place of Lemma 1.1 referenced there. □

The next theorem states the main result of this section. Let again φ_T be an arbitrary element of V_H. Then the numerical approximation of the projection coefficient $\langle y(T), \varphi_T \rangle_{\mathrm{L}^2(\Omega)}$, which is computed using the discrete operator $\Lambda^{h,\beta}$ and the projection formula (8), converges to the exact coefficient for $h, \beta \to 0$.

Theorem 3.7. *Let f^h and $y^h \chi_\omega$ be the projections of the known right hand side and measurements, respectively, onto V_h, where $h \leq H$. Further, let $\overline{\psi}_0^{h,\beta}$ be the unique solution of (23) and $\overline{\psi}^{h,\beta}$ be the corresponding solution of (17). For every $\varphi_T \in V_H$ the discrete approximation*

$$c^{h,\beta} := \left\langle f^h, \varphi^h(t; \overline{\psi}^{h,\beta} \chi_\omega) \right\rangle_{\mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))} - \left\langle y^h, \overline{\psi}^{h,\beta} \right\rangle_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))}$$

satisfies

$$\lim_{\beta\to 0} \lim_{h\to 0} c^{h,\beta} = \langle y(T), \varphi_T \rangle_{\mathrm{L}^2(\Omega)}.$$

Proof. From Theorem 3.6 and the arguments from Proposition 3.2 (specifically, (21) and (22) with $t \in [0, T]$ instead of $t = 0$), we get

$$\lim_{h\to 0} c^{h,\beta} = \left\langle f, \varphi(t; \overline{\psi}^{\beta} \chi_\omega) \right\rangle_{\mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))} - \left\langle y, \overline{\psi}^{\beta} \right\rangle_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))},$$

where $\overline{\psi}_0^{\beta}$ is the unique solution given by Proposition 2.7 and $\overline{\psi}^{\beta}$ is the corresponding solution of (11). Employing Theorems 2.7 and 2.3 for the convergence in β, we obtain

$$\lim_{\beta\to 0} \lim_{h\to 0} c^{h,\beta} = \langle f, \varphi(.; \overline{v}) \rangle_{\mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))} - \langle y, \overline{v} \rangle_{\mathrm{L}^2(0,T;\mathrm{L}^2(\omega))},$$

where $\overline{v}$ is the unique solution of Problem 2.1. Noting that this solution is by construction a null control for the original φ_T, the right hand side is equal to $\langle y(T), \varphi_T \rangle_{\mathrm{L}^2(\Omega)}$ due to Corollary 1.5. □

4 Numerical calculation of approximation

In the previous section, we have detailed an algorithm for the calculation of coefficients $\langle y(T), \varphi_T \rangle_{\mathrm{L}^2(\Omega)}$ of the projection of the unknown vector $y(T)$ onto any element

φ_T of the finite element space V_H. The next step towards a complete forward data assimilation procedure is the choice of an $L^2(\Omega)$-orthonormal basis $\{\Phi_1, \Phi_2, \ldots, \Phi_n\}$ for V_H. We are then able to compute an approximation $w \in V_H$ of $y(T)$:

$$w := \sum_{k=1}^{n} \langle y(T), \Phi_k \rangle_{L^2(\Omega)} \Phi_k. \tag{26}$$

We will address convergence of this approximation to $y(T)$ for $H \to 0$ in section 4.1.

For a practical application of this numerical method, it is crucial to use an orthonormal basis which gives a good approximation using only a small subset of basis vectors. To this end, we propose the use of proper orthogonal decomposition (POD) of a nodal basis of V_H. This approach (as opposed to the method of snapshots [23, 14], where POD is applied to an ensemble of chosen target states) has the added benefit of being independent of the measurement data; all the necessary vectors for the approximation of $y(T)$ can therefore be pre-computed independently and in parallel. On the other hand, this means that the POD basis cannot be a priori optimally adapted to specific instances of the problem; this shortcoming will be addressed in section 4.3. We conclude the section with a summary of the entire forward data assimilation algorithm.

4.1 Convergence in $H \to 0$

Let $\{\Phi_1, \Phi_2, \ldots, \Phi_n\}$ be a basis of V_H which is orthonormal with respect to the inner product of $L^2(\Omega)$. The following theorem states the convergence of the corresponding discrete approximation to $y(T)$ for $H \to 0$.

Theorem 4.1. *Let w be defined by equation (26). Then,*

$$\lim_{H \to 0} w = y(T) \tag{27}$$

holds.

Proof. This statement follows from the fact that $w = \mathcal{P}_H(y(T))$, where $\mathcal{P}_H : L^2(\Omega) \to V_H$ is the projection operator on the $n-$dimensional space V_H. By definition of the projection operator $\mathcal{P}_H$, we have

$$\langle \mathcal{P}_H(y(T)), \varphi_T \rangle_{L^2(\Omega)} = \langle y(T), \varphi_T \rangle_{L^2(\Omega)}$$

for every $\varphi_T \in V_H$. Since the vectors Φ_k form an orthonormal basis, we obtain

$$\mathcal{P}_H(y(T)) = \sum_{k=1}^{n} \langle \mathcal{P}_H(y(T)), \Phi_k \rangle_{L^2(\Omega)} \Phi_k = w. \tag{28}$$

It is a direct consequence of the approximation properties of the finite element basis that

$$\lim_{H \to 0} \| u - \mathcal{P}_H(u) \|_{L^2(\Omega)} = 0$$

for every $u \in L^2(\Omega)$. Combined with (28), this yields (27). □

4.2 Proper orthogonal decomposition

We now address the construction of the orthonormal basis $(\Phi_k)_{k=1}^n$ using proper orthogonal decomposition (also known as *principal component analysis* or *Karhunen-Loève decomposition*). For the reader's convenience, we give here a brief summary of the definition and pertinent properties of POD (cf. [12, 23], as well as [27] for a basic exposition).

The fundamental idea of POD is to approximate a set of given vectors $x_i \in \mathbb{R}^n$, $i = 1, \dots, m$ by their projection onto a small set of vectors $u_k \in \mathbb{R}^n$, $k = 1, \dots, l$ with $l \ll m$, which are additionally orthonormal with respect to a weighted inner product

$$\langle \eta, \xi \rangle_W := \langle \eta, W\xi \rangle_{\mathbb{R}^n}.$$

Here, $W \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, and $\eta, \xi \in \mathbb{R}^n$ are vectors. Since W is thus diagonalizable, we can define $W^{\frac{1}{2}}$ in the usual way and write

$$\langle \eta, \xi \rangle_W = \left\langle W^{\frac{1}{2}}\eta, W^{\frac{1}{2}}\xi \right\rangle_{\mathbb{R}^n}.$$

Furthermore, let $\|\cdot\|_W$ denote the norm induced by this inner product. In the current context, we identify the n-dimensional subspace $V_H \subset \mathrm{L}^2(\Omega)$ with $\mathbb{R}^n$ via the canonical coordinate isomorphism, and $\langle \cdot, \cdot \rangle_W$ is chosen so that it approximates $\langle \cdot, \cdot \rangle_{\mathrm{L}^2(\Omega)}$.

A POD basis is then defined as the set of l orthonormal vectors which on average best approximates the given vectors $\{x_i\}_{i=1}^m$ in the W-norm:

Definition 4.2. *Let $x_i \in \mathbb{R}^n$, $i = 1, \dots, m$ and $l \in \{1, \dots, m\}$ be given. A solution $\{u_k \in \mathbb{R}^n : k = 1, \dots, l\}$ of*

$$\min_{\{u_k\}_{k=1}^l} \sum_{i=1}^m \left\| x_i - \sum_{k=1}^l \langle x_i, u_k \rangle_W u_k \right\|_W^2 \quad \text{s.t. } \langle u_i, u_j \rangle_W = \delta_{ij}, \quad 1 \le i, j \le l$$

is called POD basis of rank l.

The construction of a POD basis can be carried out via singular value decomposition (SVD). For this, we consider the vectors x_i as columns of a matrix $X \in \mathbb{R}^{n\times m}$ having rank $r \le \min(m, n)$, and apply the SVD to

$$W^{\frac{1}{2}}X =: \overline{X} = \overline{U}\Sigma\overline{V}^T,$$

with $\overline{U} \in \mathbb{R}^{n\times n}$ and $\overline{V} \in \mathbb{R}^{m\times m}$ orthogonal, and Σ a diagonal matrix containing the singular values σ_i (assumed to be ordered by descending magnitude). We then have:

Proposition 4.3. *Let $x_i \in \mathbb{R}^n$, $i = 1, \dots, m$ denote the columns of $X \in \mathbb{R}^{n\times m}$, and $\overline{U}\Sigma\overline{V}^T$ be the SVD of $W^{\frac{1}{2}}X$. The POD basis $\{u_k\}_{k=1}^l$ of rank $l \le \operatorname{rank}(X)$ is then given by the first l columns of*

$$U := W^{-\frac{1}{2}}\overline{U}.$$

In practice, one makes use of the properties of the SVD by first calculating the vectors $\{\overline{v}_k\}_{k=1}^{l}$ as the solutions of the symmetric eigenvalue problem

$$X^T W X \overline{v}_k = \lambda_k \overline{v}_k, \qquad k = 1, \ldots, l. \tag{29}$$

A POD basis is then given by

$$u_k = \frac{1}{\sqrt{\lambda_k}} W^{-\frac{1}{2}} \overline{X} \overline{v}_k = \frac{1}{\sqrt{\lambda_k}} X \overline{v}_k, \quad k = 1, \ldots, l,$$

if the eigenvectors $\overline{v}_k$ are chosen to be normalized (with respect to the standard Euclidean norm).

To apply now POD to the problem of forward data assimilation, we need to fix the set of vectors x_i to be approximated in this way. Since we will be working with a finite element discretization of functions in the parameter space $\mathrm{L}^2(\Omega)$, a reasonable idea is to use directly the basis vectors $\{B_i\}_{i=1}^{n}$ of the corresponding ansatz space $V_H \subset \mathrm{L}^2(\Omega)$. According to the canonical coordinate isomorphism, this is equivalent to setting

$$X := I_n,$$

I_n being the $n \times n$ identity matrix.

Now, for any two vectors $\xi = \sum_{i=1}^{n} \xi_i B_i$ and $\eta = \sum_{i=1}^{n} \eta_i B_i$ in V_h, we have that

$$\langle \xi, \eta \rangle_{\mathrm{L}^2(\Omega)} = (\xi_1, \ldots, \xi_n) M (\eta_1, \ldots, \eta_n)^T,$$

with the *mass matrix* $M \in \mathbb{R}^{n \times n}$, $M = (m_{ij})_{i,j=1}^{n}$,

$$m_{ij} = \int_\Omega B_i B_j \, dx, \qquad 1 \leq i, j \leq n.$$

We thus define the weighted inner product $\langle \cdot, \cdot \rangle_W$, which approximates the $\mathrm{L}^2(\Omega)$ inner product, by setting

$$W := M.$$

With these choices, equation (29) reduces to

$$M \overline{v}_k = \lambda_k \overline{v}_k, \qquad k = 1, \ldots, l.$$

Thus, the construction of our POD basis amounts to computing the first l eigenvectors of the mass matrix, which can efficiently be performed by a sparse Krylov method (cf. section 5.1).

Hereafter, the orthonormal basis $\{\Phi_1, \Phi_2, \ldots, \Phi_n\}$ of V_H used to reconstruct $y(T)$ consists of the POD vectors introduced in this section, i.e., we set $\Phi_k := u_k$ for $k = 1, \ldots, n$.

4.3 Adaptive reordering and termination

As stated above, the POD basis so constructed depends only on the geometrical parameters of the problem, and is therefore independent of the other data, such as the coefficients or right hand side of the state equation, and especially of the measurements. While this has advantages for the practical computation, it also means that such a basis cannot be optimal for all data sets. Indeed, a basis vector corresponding to a large eigenvalue could be associated with a small projection coefficient of the unknown final state.

Since the calculation of the controls is computationally the most expensive step of the assimilation procedure, it is critical that this is only performed for basis elements likely to have a large contribution. Specifically, we are interested in a permutation $\tau : \mathbb{N} \to \mathbb{N}$ which minimizes the absolute error

$$e_a(l) := \left\| y^H(T) - w_l \right\|_M,$$

where $y^H(T) := \mathcal{P}_H(y(T))$ is the orthogonal projection of the final state onto V_H, and

$$w_l := \sum_{k=1}^{l} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M \Phi_{\tau(k)}$$

is its approximation using l reordered basis elements.

We can give a more explicit expression of the error. First we note that

$$\left\| y^H(T) \right\|_M^2 = \left\| \sum_{k=1}^{n} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M \Phi_{\tau(k)} \right\|_M^2 = \sum_{k=1}^{n} \left\langle y^H, \Phi_{\tau(k)} \right\rangle_M^2,$$

due to the orthonormality of the Φ_k. Similarly, we have that

$$\begin{aligned} e_a(l)^2 &= \left\| y^H(T) - \sum_{k=1}^{l} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M \Phi_{\tau(k)} \right\|_M^2 \\ &= \left\| y^H(T) \right\|_M^2 - \sum_{k=1}^{l} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M^2 = \sum_{k=l+1}^{n} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M^2. \end{aligned} \tag{30}$$

Proceeding inductively for $k = 1, \ldots, l$, this shows that to minimize $e_a(k)$, one should (as intuitively suggested) choose in each step that POD basis element onto which the projection coefficient of $y^H(T)$ is maximal.

Although this coefficient is of course unknown, it can be estimated using the given distributed measurement $y\chi_\omega$. Let $R : L^2(\omega) \to V_H \subset \mathrm{L}^2(\Omega)$ be a discrete interpolation operator from the subdomain ω to the whole domain Ω. While an interpolation $R(y^H(T)\chi_\omega)$ will not give a satisfactory approximation for $y^H(T)$, it can give a rough estimate of the structure of the solution, which in turn can serve as the basis for estimating the relative importance of the POD basis elements. The permutation τ is found by rearranging the basis elements according to the magnitude of their inner product

with the interpolated measurement $R(y^H(T)\chi_\omega)$ (which can be calculated without the costly solution of the corresponding control problem).

The numerical stability of the described method can be improved significantly by an iterative approach. Instead of interpolating the (projected) measurement $y^H(T)\chi_\omega$ directly, one can first subtract the already computed approximation w_l and interpolate the result $(y^H(T) - w_l)\chi_\omega$. This is due to the fact that for all $m = 1, \dots, n-l$, it holds that

$$\begin{aligned}\left\langle y^H(T) - w_l, \Phi_{\tau(l+m)} \right\rangle_M &= \left\langle y^H(T) - \sum_{k=1}^{l} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M \Phi_{\tau(k)}, \Phi_{\tau(l+m)} \right\rangle_M \\ &= \left\langle y^H(T), \Phi_{\tau(l+m)} \right\rangle_M,\end{aligned}$$

again by the orthonormality of the POD basis. Hence, instead of an initial reordering of the basis, in each step the next POD basis element is chosen based on the (interpolated) measurement and the already computed approximation. The reason for the improved accuracy of $\left\langle R[(y^H(T) - w_l)\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M$ as an estimate of $\left\langle y^H(T), \Phi_{\tau(l+m)} \right\rangle_M$ compared to $\left\langle R[y^H(T)\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M$ can easily be seen when the interpolation operator R is linear. In this case, the interpolation error projected onto the basis vector $\Phi_{\tau(l+m)}$ can be expressed as

$$\begin{aligned}e_{int}(l+m) &= \left\langle y^H(T), \Phi_{\tau(l+m)} \right\rangle_M - \left\langle R[y^H(T)\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M \\ &= \left\langle y^H(T) - R[(y^H(T) - w_l)\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M + \left\langle R[w_l\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M \\ &= \left\langle (y^H(T) - w_l) - R[(y^H(T) - w_l)\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M \\ &\quad + \sum_{k=1}^{l} \left\langle y^H(T), \Phi_{\tau(k)} \right\rangle_M \left\langle R[\Phi_{\tau(k)}\chi_\omega], \Phi_{\tau(l+m)} \right\rangle_M\end{aligned}$$

The error $e_{int}(l+m)$ therefore consists of two parts: the above described interpolation error of the difference of the measurement and the current approximation, and a contribution from the previously computed coefficients. Now while the first term will be small for large l since $\left\| y^H(T) - w_l \right\|_M \to 0$ for $l \to n$, this is not true in general for the second term, which might even increase with l and therefore dominate the error when working with the initial interpolation $R(y^H(T))$ only. Using only the first term will therefore yield a much sharper estimate later in the iteration (i.e., for larger l).

Remark 4.4. *Given an estimate of $\left\| y^H(T) \right\|_M$—e.g., from interpolation—and the previously computed projections, it is possible to estimate the error $e_a(l)$ using relation (30). This can be used as an effective and easily evaluated stopping criterion for the numerical assimilation procedure.*

Algorithm 2 shows how the ingredients of the forward data assimilation process are combined. This comprises the computation of the POD basis, reordering of these components, and the solution of the control problems.

Algorithm 2 Forward data assimilation

Calculate POD decomposition $\Phi_1, \dots, \Phi_l$ of mass matrix
initialize approximation $w_0 \leftarrow 0$
for $k = 1, 2, \dots$ **do**
  compute interpolant $\widetilde{y}_T := R((y^H(T) - w_{k-1})\chi_\omega)$
  for all remaining POD components Φ_j, $j \notin \{\tau(1), \dots, \tau(k-1)\}$ **do**
    compute estimated projection coefficient $\widetilde{c}_j := \left\langle \widetilde{y}_T, \Phi_j \right\rangle_M$
  end for
  choose POD component $\Phi_{\tau(k)}$ with largest $\widetilde{c}_{\tau(k)}$
  solve Problem 2.2: Set $\varphi_T = \Phi_{\tau(k)}$ and compute control v and solution φ using Algorithm 1
  compute corresponding projection coefficient $c_k \approx \left\langle y^H(T), \varphi \right\rangle$ using (8)
  update approximation $w_k \leftarrow w_{k-1} + c_k \Phi_{\tau(k)}$
  estimate error $\varepsilon_{rel} = \frac{\|w_k\|_M}{\|y^H(T) - w_k\|_M}$
  if $\varepsilon_{rel} < tol$ **then**
    return w_k
  end if
end for

5 Numerical experiments

In this section, we give some details on the implementation of Algorithm 2, and present numerical results for test problems in two and three spatial dimensions.

5.1 Implementation

The method described above is implemented in C++ within the finite element framework *deal.II* [2]. Most calculations are performed on an Intel quad core workstation with 2.4 GHz and 2 GB of RAM. The computational work can be distributed to an arbitrary number of threads, since the controls for every POD component can be calculated independently. Each thread performs the selection of the next basis element in a critical (synchronized) section of the program, using for the reordering heuristic the combined approximation computed from the contributions of all threads so far. The algorithm is parallelized using the OpenMP application programming interface. The more time consuming calculations with $\Omega \subset \mathbb{R}^3$ are performed on a shared memory system with 16 Opteron CPUs at 2.8 GHz and 64 GB of RAM.

Discretization We employ the discretization scheme introduced in section 3.1, using bilinear finite elements ($s = 1$) on a quadrilateral mesh for the discretization in space and the discontinuous Galerkin method of order $r = 1$ in time (which in our tests gave significantly better results than $r = 0$). All integrals for the assembly of the finite element matrices are computed using Gauß quadrature with 2 points per dimension, which is equivalent to exact integration in this setting. The evaluation of inner

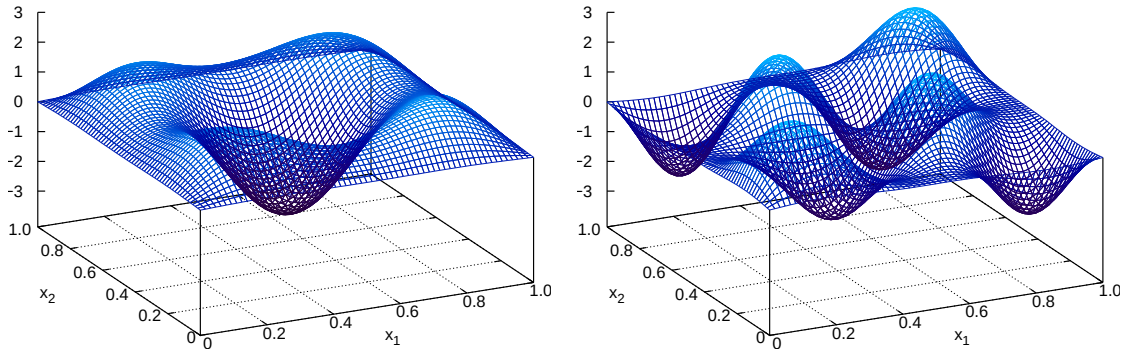

Figure 1: Two POD basis element for a piecewise linear nodal basis on the unit square (mesh size $h = \frac{1}{64}$).

products $\langle \cdot, \cdot \rangle_{L^2(0,T;L^2(\Omega))}$ is performed by first applying Gauß quadrature in space and then integrating in time using the trapezoidal rule. This amounts to exact integration for elements of S_1 and S_1^*, since these are piecewise linear functions in time. The linear systems in each time step are solved using a direct sparse solver from UMFPACK. (For higher dimensional systems, iterative methods like GMRES are used instead.)

POD BASIS The computation of the reduced POD basis of rank l entails the calculation of the l largest magnitude eigenvectors of the mass matrix M. There are several well known efficient algorithms to solve eigenvector problems for high dimensional sparse matrices, such as the methods by Lanczos, Arnoldi and Jacob-Davidson. Since the matrix M is symmetric and positive definite, and we are only interested in eigenvalues near the boundary of the spectrum, we use the implicitly restarted Lanczos method [24]. Deflation techniques of locking and purging are implemented to improve the convergence to the desired eigenvalues. We refer the reader to [16] for details on these techniques. Figure 1 shows two POD basis elements computed from the full set of piecewise linear nodal basis elements for a uniform mesh of size $h = \frac{1}{64}$ on the unit square.

ERROR ESTIMATION AND BASIS REORDERING As already mentioned, a data specific reordering of the POD basis is critical to the performance of the method. Linear interpolation of the measured data $y(T)\chi_\omega$ is an effective and computationally inexpensive way to construct an estimate of $y^H(T)$, which is sufficient for the purpose of improving the ordering.

The interpolation of the measurement data (and boundary conditions) is performed by constructing a Delaunay triangulation $\mathcal{T}$ of the (discretized) measurement subdomain ω and the boundary Γ, where the solution is known on the corners of every element. For each point $p \in \Omega$ there exists at least one triangle $T(p) \in \mathcal{T}$ containing p. The barycentric coordinates of p with respect to $T(p)$ are used for linear interpolation to construct the interpolant $\tilde{y}(T)$. For these geometric calculations, the C++ library

CGAL [1] is employed.

Although the primary focus of this heuristic is the rearrangement of the POD basis, it can serve as an estimate of $\|y^H(T)\|$ at the same time. Due to equation (30), this is equivalent to estimating the absolute error $e_a(l)$. We therefore have a stopping criterion at no further computational cost. However, one drawback of the linear interpolation is its sensitivity to errors in the measured data. The influence of such noise on the numerical assimilation procedure is investigated at the end of section 5.2.

Remark 5.1. *There are of course several other possible heuristics for error estimation. One simple idea is to compute the difference of measurements and approximation on the subset ω and approximate the absolute error by $e_a(l) \approx \|(y(T) - w_l)\chi_\omega\|$. While this does not give us any information about the coefficients, it is a competitive stopping criterion. Another method is the approximation of $\partial_t y(t)$ at time $t = T$. This can be achieved by interpolation of the difference quotient $(y(T) - y(T - \Delta t))\chi_\omega / \Delta t$, where Δt is the time step length. Once we have an estimate of the time derivative, we can get the corresponding estimate for $y(T)$ by fixing $t = T$ and solving the (now stationary) boundary value problem (1).*

5.2 Two-dimensional test problem

For our two dimensional test problem, we specify as the domain the unit square $\Omega = [0,1]^2$ and the time interval $[0,1]$, i.e. $T = 1$. Since there are few known analytical solutions to the convection-diffusion equation (1) which we could use for comparisons with our results, we choose an initial value $y(0)$ and compute the corresponding numerical solution in $\Omega \times [0,1]$ on a very fine mesh. This highly accurate solution is used to generate measurements $y\chi_\omega$ for $t \in (0,T]$. (The value $y(0)$ itself must of course not be used in the assimilation algorithm.) We refer to it as the "exact solution" for the rest of this section.

The data of the state equation is chosen as follows. For simplicity, the diffusion tensor is taken as a constant scalar $c(x,t) \equiv 0.1$. A constant flux $b(x,t) \equiv (1,1)^T$ for all $t \in [0,T]$ is specified. As the right hand side f of (1), we choose a half-ellipsoid with time dependent length:

$$f(x,t) = \begin{cases} 10\cos(3\pi t)\sqrt{r^2 - \|x-p\|_2^2}, & \|x-p\|_2 \le r, \\ 0, \quad \text{otherwise}, \end{cases} \tag{31}$$

with radius $r = 0.2$ and center $p = (0.5, 0.5)^T$. The initial value $y(0)$ used to generate the exact solution is a combination of sine terms:

$$[y(0)](x_1,x_2) = 10\sin(3x_1\pi)[\sin(2x_2\pi) + \sin(3x_2\pi) + \sin(4x_2\pi)].$$

The parameters in the numerical assimilation algorithm are set in the following way. The subdomain ω consists of 49 circles with radius $r = \frac{1}{42}$ each, which are arranged in a regular 7×7 grid. Together, they cover 8.7% of the area of Ω. We use a mesh size of $h = \frac{1}{128}$ in space and a time step length of $\Delta t = \frac{1}{256}$.

relative error	# components	# iterations	time[min]
0.10	9	74	2.4
0.05	21	166	4.2
0.02	36	286	6.4
0.01	63	412	8.9
0.005	117	515	12

Table 1: Number of POD components, total number of BiCGstab iterations and time in minutes needed to reach a given relative error level.

The control algorithm 1 is implemented using the BiCGstab method (cf. Remark 3.5), for which the relative tolerance is set to $tol_{\text{rel}} = 0.01$, using the stopping criterion

$$\|g^{(k)}\|_{\mathrm{L}^2(\Omega)} \leq tol_{\text{abs}} := \max\{tol_{\text{rel}} \cdot \|\varphi_0\|_{\mathrm{L}^2(\Omega)}, l_{\text{tol}}\}.$$

To avoid too many iterations for small values of the target norm $\|\varphi_0\|_{\mathrm{L}^2(\Omega)}$, an additional lower bound is posed on the absolute tolerance: $tol_{\text{abs}} \geq l_{\text{tol}} = 10^{-6}$. The effect is similar to a restriction on the maximal number of iterations. The parameter β is set to $\beta = 10^{-6}$, since we obtained identical results in our tests for lower values.

Let w_l be the computed approximation after l POD components and $\mathcal{I}_H(y(T))$ be the interpolation of the exact solution $y(T)$ in our finite element subspace V_H. Since the interpolation error is very small compared to the error of assimilation at the considered mesh size, we define the relative error as

$$e_{\text{rel}} := \frac{\|w_l - \mathcal{I}_H(y(T))\|_{\mathrm{L}^2(\Omega)}}{\|\mathcal{I}_H(y(T))\|_{\mathrm{L}^2(\Omega)}}.$$

We use e_{rel} as measurement of the accuracy of all numerical solutions.

The full POD basis for the given mesh size contains more than 4000 elements. However, a very small number of them is sufficient to get a close approximation of the solution. Knowing this, we include only the first 800 POD components in the reordering heuristic.

We then compute the approximation of $y(T)$ using Algorithm 2 for this set of POD components. The number of POD components needed to reach a given relative error is shown in Table 1. We also quote the total number of iterations of the BiCGstab method used to compute all corresponding controls. Note that every such iteration involves two evaluations of the discretized operator Λ^h and thus four solutions of a partial differential equation. The computational time given in the table is the time needed for the complete forward assimilation algorithm 2. This includes about 40 seconds for the computation of the POD basis, which is small compared to the assimilation itself.

Figure 2 shows a cut through the exact solution at $x_2 = 0.5$ and the corresponding approximations using 10, 25 and 75 components with relative errors of 0.091, 0.038 and 0.007, respectively.

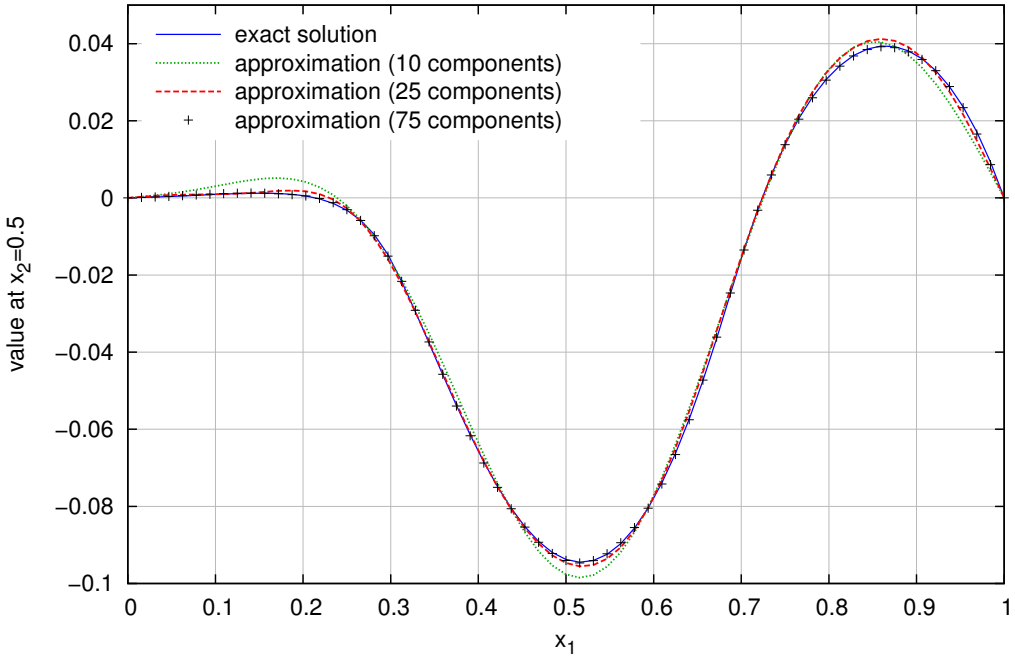

Figure 2: Cut through exact solution and approximations using 10, 25 and 75 POD components.

The development of the error versus the number of basis elements is shown in Figure 3. When no problem specific reordering of the basis is performed, some of the components with large projection coefficients (and thus large impact on the approximation) are used later in the assimilation process. Thus one can observe rather flat sections in the error curve, followed by steep steps downwards. If we use the heuristic detailed in section 5.1, we see on the other hand an almost strictly monotone decline of the contributions. In this case, about 20 components are necessary to reach a relative error of 0.05, which is otherwise not reached until the 40^{th} component. For comparison, the error from a simple interpolation of the measurement data is also included in Figure 3. Due to the very regular distribution of the measurement domain ω in our test problem, the error using interpolation only is at the relatively low level of 0.10. However, the performance of interpolation compared to the assimilation gets worse when we consider data subject to measurement errors, which is of course always the case in practice.

The same holds true when ω is smaller and its distribution less regular. Figure 4 shows the development of e_{rel} for different sizes of ω, determined by the radius r of each of the 49 circles constituting ω. Note that the basis reordering heuristic is less effective for smaller r, and that the error of pure interpolation is in fact increasing to 0.16 for the smallest radius.

In order to study the influence of measurement errors, noise is added to the measurements $y\chi_\omega$. In every time interval $I_m = (t_{m-1}, t_m]$, the reference solution is given

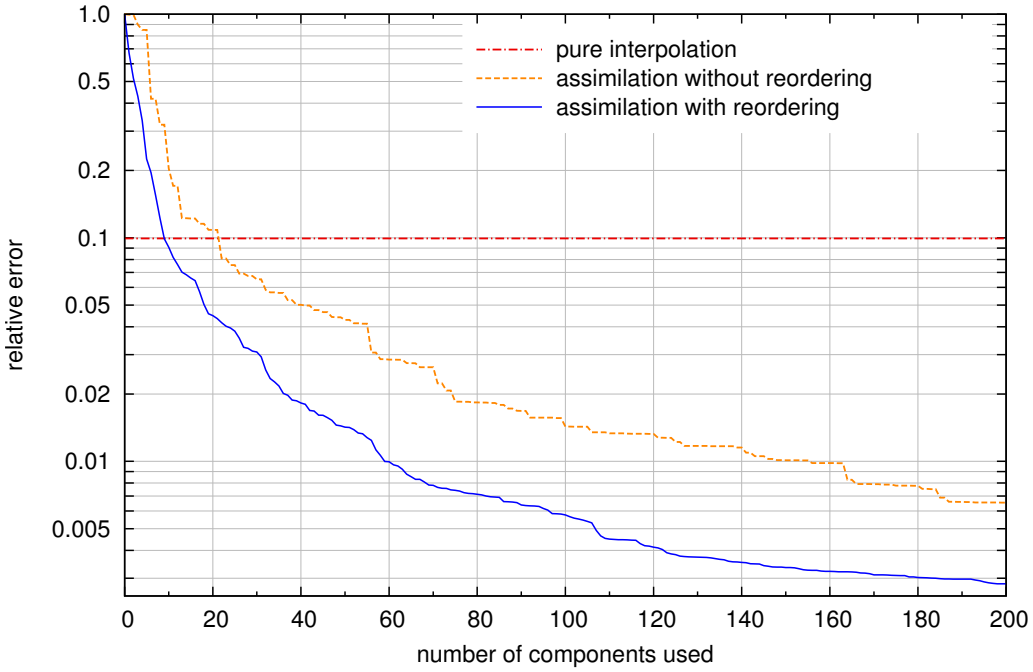

Figure 3: Relative error versus number of POD components with and without basis reordering.

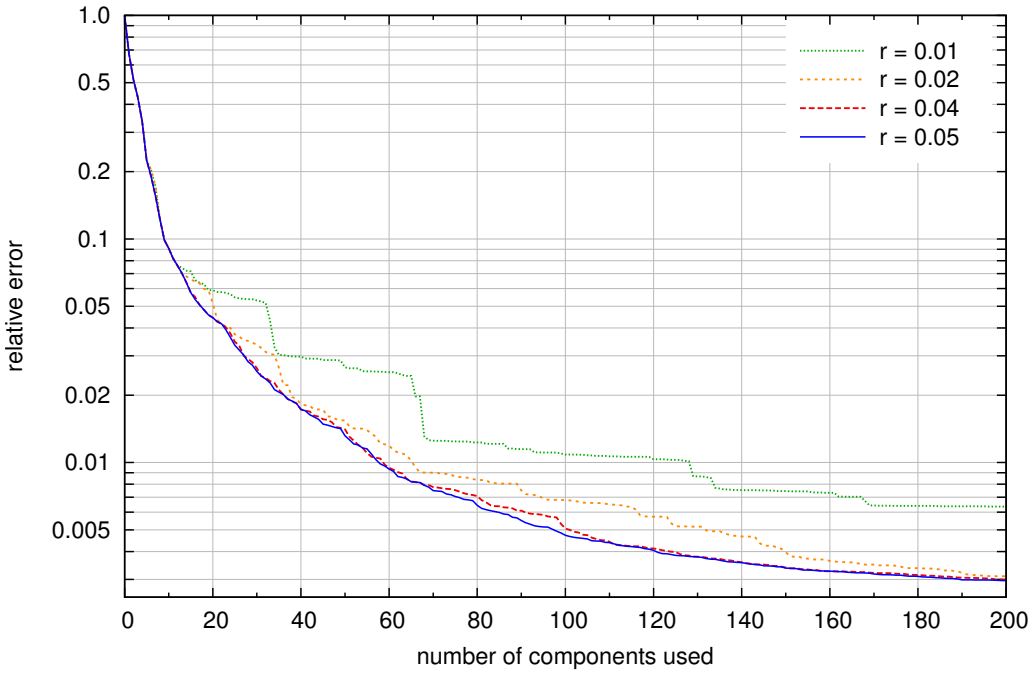

Figure 4: Relative error for different radii r of the circles in the subdomain ω.

δ	0.005	0.01	0.02	0.03	0.04	0.05
e^{δ}_{rel}	$7.1 \cdot 10^{-7}$	$4.6 \cdot 10^{-6}$	$2.0 \cdot 10^{-5}$	$4.6 \cdot 10^{-5}$	$8.1 \cdot 10^{-5}$	$1.2 \cdot 10^{-4}$

δ	0.06	0.07	0.08	0.09	0.1
e^{δ}_{rel}	$1.8 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$3.0 \cdot 10^{-4}$	$3.8 \cdot 10^{-4}$	$4.5 \cdot 10^{-4}$

Table 2: Additional error e^{δ}_{rel} caused by different noise levels δ using 200 basis elements in fixed order. Shown are averages over 100 experiments.

by a linear function in time $y(t) = y^0_m + \tau y^1_m$, where $\tau := \frac{t - t_{m-1}}{\Delta t}$. The vectors y^j_m, $j = 0, 1$, are now perturbed by adding uniformly distributed random vectors ξ^j_m. The noisy measurements

$$\widetilde{y}\chi_\omega(t) := (y^0_m + \xi^0_m) + \tau(y^1_m + \xi^1_m), \qquad t \in I_m,$$

are then used for assimilation. We define the noise level in each time interval I_m by

$$\delta_m := \frac{\|\xi^0_m\|}{\|y^0_m \chi_\omega\|} = \frac{\|\xi^1_m\|}{\|y^1_m \chi_\omega\|}.$$

In our experiments, this level is constant for all time steps: $\delta_m \equiv \delta \in [0, 1]$. This is achieved by first generating random vectors $\zeta^0_m, \zeta^1_m \in [-1, 1]^{n_\omega}$ (where n_ω is the number of grid points in ω), and then scaling to

$$\xi^j_m = \delta \frac{\|y^j_m \chi_\omega\|}{\|\zeta^j_m\|} \zeta^j_m, \quad j = 0, 1.$$

The results for different values of δ are shown in Figure 5. The assimilation algorithm proves to be quite robust with respect to errors in the measurements. This can be attributed partly to the smoothing effects of diffusion equations, and partly to the integration in the projection formula (8).

Since we are interested in the influence of noisy data on the reordering, we first look at the results of a single assimilation experiment, without averaging. Higher noise leads to less accurate predictions and therefore less strictly declining errors. In fact, this effect is responsible for the major part of the additional error caused by noisy data. To investigate the influence of noise on the forward assimilation method itself, without heuristic, a second series of experiments using a fixed order of the basis elements is performed. Naturally, the achievable accuracy is limited by the amount of noise. Splitting the total relative error into $e_{\text{rel}} = e^0_{\text{rel}} + e^{\delta}_{\text{rel}}$, where e^0_{rel} is the error without any noise and e^{δ}_{rel} is the additional error caused by noisy data, we observe $e^{\delta}_{\text{rel}} = \mathcal{O}(\delta^2)$ for the latter. Table 2 shows the results after 200 basis elements, where each entry is the arithmetic mean from 100 experiments to average the impact of randomness.

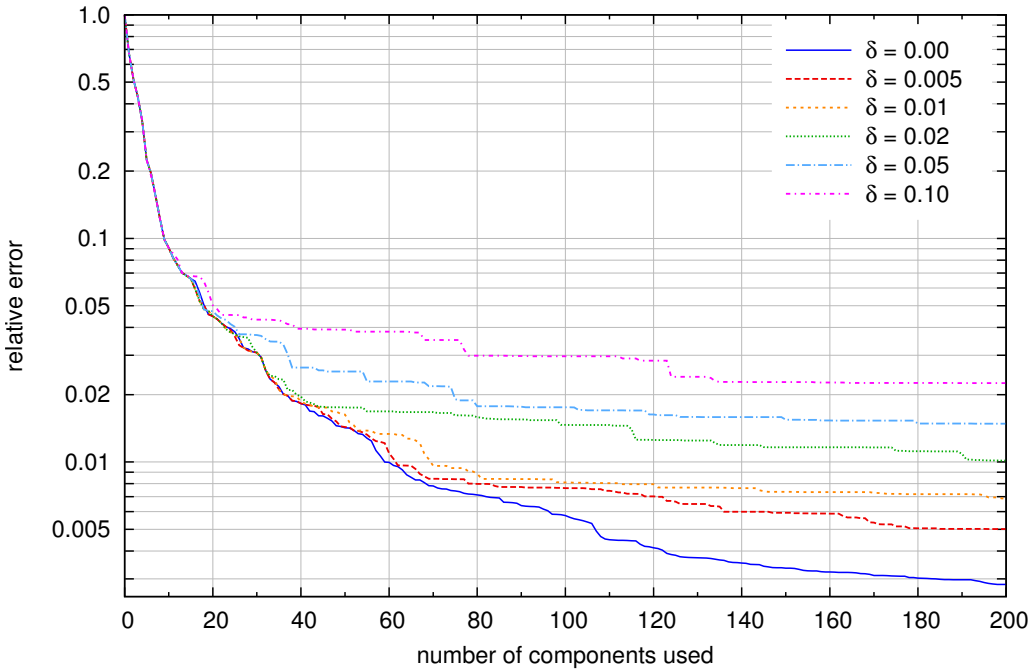

Figure 5: Relative error at different noise levels δ.

5.3 Three-dimensional test problem

To demonstrate the effectiveness of the algorithm in more than two dimensions, we extend the test problem from the previous section to $\Omega = [0,1]^3$. The diffusion coefficient $c \equiv 0.1$ and the time interval $[0,T] = [0,1]$ remain unchanged. The flux is $b(x,t) \equiv (1,1,1)^T$, and the right hand side of equation (1) is still given by (31), where $p = (0.5, 0.5, 0.5)^T$ is again the center of Ω, and the norm is now taken in $\mathbb{R}^3$. The initial value for y has an additional factor for the third coordinate, yielding

$$[y(0)](x) = 10\sin(3x_1\pi)[\sin(2x_2\pi) + \sin(3x_2\pi) + \sin(4x_2\pi)] \\ \times\ [\sin(2x_3\pi) + \sin(3x_3\pi) + \sin(4x_3\pi)].$$

The subdomain ω consists of 7^3 balls with radius $\frac{1}{42}$ whose centers are again distributed in a regular grid. Thus, the measurements now cover only 1.9% of the volume in Ω, which is far less than in our two dimensional experiments. In order to keep the computational time at a reasonable level, the mesh size is reduced to $h = \Delta t = \frac{1}{32}$. Therefore, the relative error stagnates at a higher level than before. The development of e_{rel} is displayed in Figure 6. The computational time for the assimilation is given in Table 3, which includes about 8.8 minutes needed for the POD.

relative error	# components	# iterations	time[h]
0.10	20	150	9.7
0.05	36	207	13
0.02	119	347	21

Table 3: Number of POD components, total number of BiCGstab iterations and time in hours needed to reach a given relative error level for the three-dimensional test problem.

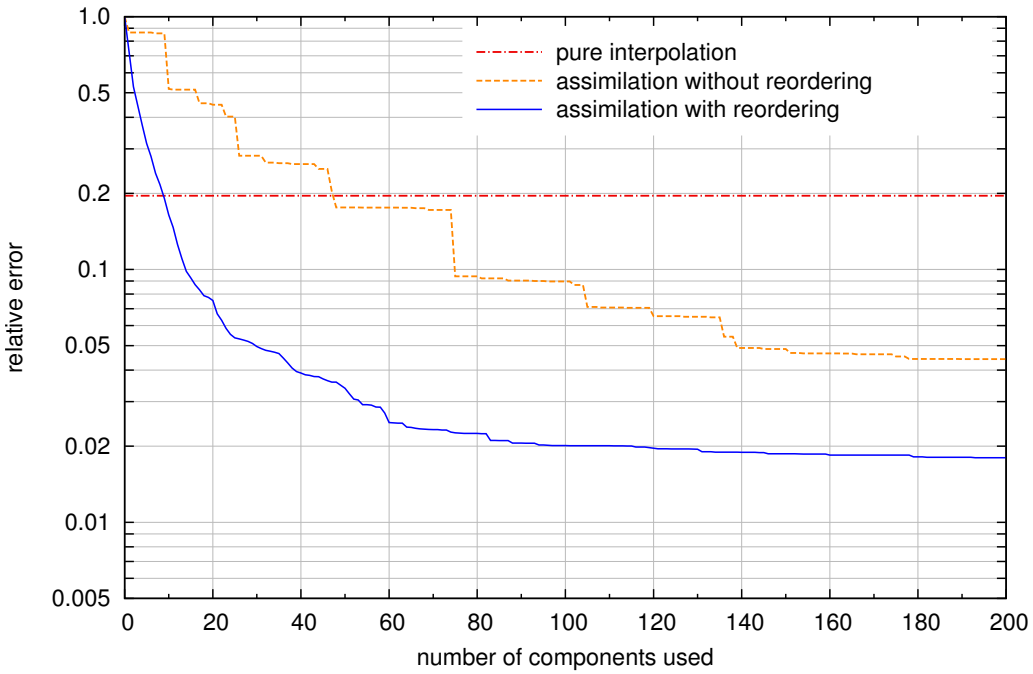

Figure 6: Relative error in three dimensional test problem.

6 Conclusion

We have presented an efficient numerical method for solving the "variational forward assimilation" problem for a linear scalar convection diffusion equation, in which the ill-posed optimal control problem is replaced with a well-posed series of exact control problems. Using a combination of proper orthogonal decomposition and adaptive basis reordering, this problem can be solved efficiently and stably, with the most time intensive calculations being pre-computable in parallel. Its main advantage, apart from computational efficiency, over optimal control methods (such as 4DVAR) consists in the absence of regularization parameters which have to be chosen dependent on the measurement error.

Future work will be concerned with more complicated equations such as the linearized Navier-Stokes equations, as well as the investigation of problems dealing with

boundary observations (which are relevant, e.g., in geophysics). Although the described approach is directly applicable in the former case, the latter needs different error estimators for the basis reordering. One possibility would be to use the solution of a stationary boundary value problem to approximate the unknown final state. Ultimately, the proposed method should be applied for the nonlinear systems relevant in the application areas mentioned in the introduction, for instance using appropriate linearization strategies.

Also of interest would be extending the proper orthogonal decomposition of the mass matrix to adaptive finite element methods, which would possibly lead to an even better adapted reduced basis.

Acknowledgments

The authors wish to thank Jean-Pierre Puel for helpful discussions, and the referees for their useful remarks. The work of Peter Hepperger was supported by by the International Graduate School of Science and Engineering (IGSSE) of Technische Universität München.

References

[1] *CGAL, Computational Geometry Algorithms Library*. URL: http://www.cgal.org.

[2] W. Bangerth, R. Hartmann, and G. Kanschat, *deal.II—a general-purpose object-oriented finite element library*, ACM Trans. Math. Softw., 33 (2007), p. 24.

[3] A. F. Bennett, *Inverse methods in physical oceanography*, Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press, Cambridge, 1992.

[4] ——, *Inverse modeling of the ocean and atmosphere*, Cambridge University Press, Cambridge, 2002.

[5] A. F. Bennett, B. S. Chua, and L. M. Leslie, *Generalized inversion of a global numerical weather prediction model*, Meteorology and Atmospheric Physics, 60 (1996), pp. 165–178.

[6] H.-P. Bunge, M. A. Richards, and J. R. Baumgardner, *Mantle-circulation models with sequential data assimilation: inferring present-day mantle structure from plate-motion histories*, Phil. Trans. Roy. Soc. A, 360 (2002), pp. 2545–2567.

[7] K. Chrysafinos and N. J. Walkington, *Error estimates for the discontinuous Galerkin methods for parabolic equations*, SIAM J. Numer. Anal., 44 (2006), pp. 349–366 (electronic).

[8] L. C. Evans, *Partial differential equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 1998.

[9] G. Evensen, *Data assimilation. The ensemble Kalman filter*, Springer-Verlag, Berlin, 2007.

[10] E. Fernández-Cara, M. González-Burgos, S. Guerrero, and J.-P. Puel, *Exact controllability to the trajectories of the heat equation with Fourier boundary conditions: the semilinear case*, ESAIM Control Optim. Calc. Var., 12 (2006), pp. 466–483 (electronic).

[11] R. Glowinski and J.-L. Lions, *Exact and approximate controllability for distributed parameter systems*, Acta Numerica, 3 (1994), pp. 269–378.

[12] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge Monographs on Mechanics, Cambridge University Press, Cambridge, 1996.

[13] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, 2002.

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

[15] F.-X. Le Dimet and O. Talagrand, *Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*, Tellus Series A, 38 (1986), pp. 97–+.

[16] R. Lehoucq and D. Sorensen, *Implicitly restarted Lanczos method*, in Templates for the solution of algebraic eigenvalue problems: A practical guide, Bai, Z.; Demmel, J.; Dongarra, J.; Ruhe, A.; Vorst, H. van d., ed., SIAM, 2000. URL: http://www.cs.utk.edu/ dongarra/etemplates/book.html.

[17] J.-L. Lions, *Exact controllability, stabilization and perturbations for distributed systems*, SIAM Rev., 30 (1988), pp. 1–68.

[18] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. II*, Springer-Verlag, New York, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 182.

[19] J.-P. Puel, *Une approche non classique d'un problème d'assimilation de données*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 161–166.

[20] ——, *A nonstandard approach to a data assimilation problem and tychonov regularization revisited*, SIAM Journal on Control and Optimization, 48 (2009), pp. 1089–1111.

[21] R. T. Rockafellar, *Duality and stability in extremum problems involving convex functions*, Pacific Journal of Mathematics, 21 (1967), pp. 167–187.

[22] D. Rozier, F. Birol, E. Cosme, P. Brasseur, J. M. Brankart, and J. Verron, *A reduced-order Kalman filter for data assimilation in physical oceanography*, SIAM Rev., 49 (2007), pp. 449–465 (electronic).

[23] L. Sirovich, *Turbulence and the dynamics of coherent structures. I–III. Coherent structures*, Quart. Appl. Math., 45 (1987), pp. 561–590.

[24] D. C. Sorensen, *Numerical methods for large eigenvalue problems.*, Acta Numerica, 11 (2002), pp. 519–584.

[25] O. Talagrand and P. Courtier, *Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory*, Quarterly Journal of the Royal Meteorological Society, 113 (1987), pp. 1311–1328.

[26] V. Thomée, *Galerkin finite element methods for parabolic problems*, vol. 25 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 2006.

[27] S. Volkwein, *Model reduction using proper orthogonal decomposition*. Lecture notes, URL: http://www.uni-graz.at/imawww/volkwein/POD.pdf, 2008.

[28] J. Wloka, *Partial differential equations*, Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.