# A REDUCED BASIS KALMAN FILTER FOR PARAMETRIZED PARTIAL DIFFERENTIAL EQUATIONS *

Markus Dihlmann[1] and Bernard Haasdonk[1]

**Abstract.** The Kalman filter is a widely known tool in control theory for estimating the state of a linear system disturbed by noise. However, when applying the Kalman filter on systems described by parametrerized partial differential equations (PPDEs) the calculation of state estimates can take an excessive amount of time and real-time state estimation may be infeasible. In this work we derive a low dimensional representation of a parameter dependent Kalman filter for PPDEs *via* the reduced basis method. Thereby rapid state estimation, and in particular the rapid estimation of a linear output of interest, will be feasible. We will also derive *a posteriori* error bounds for evaluating the quality of the output estimations. Furthermore we will show how to verify the stability of the filter using an observability condition. We will demonstrate the performance of the reduced order Kalman filter and the error bounds with a numerical example modeling the heat transfer in a plate.

## 1. Introduction

When using a mathematical model to describe and predict the behavior of a system – be it of technical, physical, biological or economic nature – we might make the experience that our predictions deviate considerably from that what happens in the real world system. Although having elaborated this mathematical model to our best knowledge, capturing the behavior of a phenomenon with a mathematical representation in every detail is nearly impossible. These modeling errors can ultimately lead to a deviation between the model and the real world system. Another source for errors are randomly occurring disturbances from the environment affecting the system. Due to their random nature, they are impossible to predict and can not be included into the model. An example is the modeling of weather phenomena, where it is obviously impossible to take into account every detail on a global scale and clearly this system is subject to unpredictable disturbances.

For some applications it is necessary to fill this gap of information between the predictions of a mathematical model and the actual state of the system. The term "state" stands here for the entity of variables providing a complete representation of the internal condition of the system. Note, that it is often infeasible or physically impossible to measure the entire state of a system. However, it can be crucial to have knowledge about the system's state either for reasons of surveillance or for using the information to control the system. The process of estimating the actual state of a system using measurement data is known as data assimilation.

The methods for data assimilation in evolution problems can be grouped into variational methods and filtering. While in [16, 37, 40] variational methods are treated, we consider in this article the filtering. As we will focus on linear time invariant systems being subject to normally distributed white noise, the Kalman filter – first introduced in [32] – provides optimal state estimates. A brief historical overview over the development towards this filter can be found in ([50], p. 485*ff.*). The Kalman filter is a well known and widely used tool in control theory and plays a role in a large range of engineering applications, *cf.* [51] for examples. While the filter was originally formulated in state space description for finite dimensional dynamical systems, there exist also studies deriving this optimal filter for partial differential equations (PDEs). In particular, optimal filters were derived in time-continuous and infinite dimensional analytic formulation [14, 15, 21, 24] and weak formulation [2, 24]. Here, we will derive a time-discrete infinite dimensional Kalman filter in weak formulation omitting semigroup theory.

However, when applying the Kalman filter to discretized PDEs, we face the difficulty of excessively large computational costs for performing state estimations. This is due to the fact, that the discretization of PDEs often results in very high dimensional dynamical systems and that the Kalman filter estimations require the integration of an error covariance system with dimension equal to the square of the disturbed system's dimension. Consequently, it is often infeasible to perform real-time state estimations for systems described by PDEs.

To solve this problem we will derive a low dimensional Kalman filter using the reduced basis method. With this Reduced Basis Kalman Filter (RBKF) rapid state estimations for linear parametrized partial differential equations (PPDEs) will be possible. The reduced basis (RB) method is a model reduction technique particularly suited for parameterized problems. An introduction can be found in [49] and RB for time dependent problems is treated in [25, 28] for example. The reduced order Kalman filter estimation procedure will be divided into two phases. In a preparative offline phase, the reduced basis is constructed and parameter independent operator components are projected onto a low dimensional space. In the subsequent online phase, rapid state estimations for given system parameters can be performed.

An overview of existing work on the topic of reduced order Kalman filters, also known as suboptimal filtering schemes, can be found in [50, 52]. Most suboptimal filters can be roughly grouped in two categories: either a dimension reduction of the model is performed and subsequently a Kalman filter for that reduced order model is derived (*e.g.* [22, 45]) or a reduced rank approximation of the error covariance propagation is used in combination with a full dimensional state propagation and update (*e.g.* [9, 13, 33, 52, 54]). A more refined categorization can be found in [52].

One could argue, that model reduction of the Kalman filter was not necessary, as the stationary Kalman gain could be calculated beforehand and then be used later for updates in the state estimations, *cf.* [44, 54]. Many systems in engineering, however, are parameter dependent. Those parameters might change from situation to situation and then the pre-calculated Kalman gain could be useless and even lead to an unstable filter, meaning that the estimation error variance is unbounded for infinite time. Thus, we will derive here a parameter dependent reduced order Kalman filter to allow rapid state estimations for parametrized systems with random disturbances.

The Kalman filter does not only give state estimations, but it also provides the estimation error covariances and is thereby giving an information about the accuracy and the reliability of the estimates. While this information is missing in other reduced order filter approaches, we will present here rigorous upper bounds for the expectation and variance of the estimation error for linear outputs. Thereby the quality of the RBKF estimations can be quantified. This verification is of importance as the use of suboptimal filters can possibly lead to unstable filters, *cf.* [52]. See also [36] studying the effects of model errors on the Kalman filter considering a wave equation. Note, that in [23] similar error bounds were derived, yet, calculating these bounds is computationally

expensive, while the error bounds derived here will be rapidly computable online. To be more specific, the error estimator will have a computational complexity polynomial in the reduced dimension and independent of the dimension of the original PDE discretization. Furthermore, we will show that the stability of the RBKF can be verified *a priori via* an observability condition. Note, that our error bounds will be given with respect to a very high dimensional solution approximation, which we will call the "true" solution. In contrast to our error bounds, the author in [1] derives bounds for the variance of the error between the infinite dimensional and the discretized Kalman filter at infinite time. Besides error bounds for the RBKF, we will also obtain approximation error bounds for linear reduced basis models being subject to noise. Compared to [4, 5, 29] we are considering in the present work a less general problem setting as we restrict the randomness to a stochastic forcing of the PPDE with Gaussian white noise. On the other hand, we are able to give, through this restriction, an ad-hoc information on error bounds during the online simulations without the need to perform a sampling or to perform Monte-Carlo simulations.

While in other projection based reduced order filters, the influence of the noise is not taken into account when choosing the low dimensional subspace, we will propose a new method for building up a reduced basis space taking into account information about the assumed stochastic forcing. Thereby, better approximations – especially considering the error variance – can be achieved.

Recently, there have been advances in combining model reduction and other data assimilation techniques. In [40] data assimilation for stationary problems *via* a variational formulation is proposed and an error analysis is presented. In [39, 48] the empirical interpolation method is used for data assimilation and for parameter estimation [48]. A reduced order Unscented Kalman filter is proposed in [42] demonstrating also the possibility to estimate parameter values with reduced order filters. In [19] statistical modeling is used to improve error bounds for reduced order models.

The new contributions in this article are the following: (I) we treat parameter dependent PDEs and derive a parameter dependent reduced order Kalman filter using the reduced basis method; (II) we propose error bounds for the stochastic moments of the output estimates; (III) we introduce a new reduced basis generation procedure suited for linear systems corrupted by white noise.

The article is structured as follows. In Section 2 the problem setting and the fundamental definitions are presented. Also the time discrete infinite dimensional Kalman filter in weak formulation is derived. This formulation is the basis for the derivation of the RBKF in Section 3. In Section 4 a detailed error analysis is presented. The numerical experiments in Section 5 illustrate the procedure of rapid state estimation with the RBKF and show the performance. Finally, we summarize and conclude in Section 6.

## 2. Problem setting

In the following we are interested in solutions to a parametrized parabolic partial differential equation on a domain $\mathcal{D} \subset \mathbb{R}^d$ with $d \in \mathbb{N}$ on a time interval $[0, T]$, $T > 0$. The domain $\mathcal{D}$ has a Lipschitz continuous boundary $\Gamma$. The time interval is divided into $K$ intervals of equal length $\Delta t = \frac{T}{K}$ and at their intersections we define the time steps $t^k = k\Delta t$ with $k = 0, \ldots, K$. The set of all time steps is $\mathbb{T} = \{t^k | k = 0, \ldots K\}$.

Furthermore, we assume $Z = Z(\mathcal{D})$ to be a real separable Hilbert space of functions over the domain $\mathcal{D}$ with inner product $(\cdot, \cdot)$ and norm $\|\cdot\|$.

As the considered parametrized PDE will be stochastically forced, we consider $(\Omega, \mathcal{F}, \mathrm{Pr})$ to be a complete probability space with the sure event $\Omega$, the non-empty Borel field of events $\mathcal{F}$ and $\mathrm{Pr}$ a probability measure on $\mathcal{F}$. Then we use the standard definitions for random functions and random processes (see [15,34]): A $Z$-valued random function is a map $z : \Omega \to Z$ which is measurable with respect to the Pr-measure. If a random function $z$ is integrable ($z \in L_1(\Omega, \mathrm{Pr}; Z)$ $cf.$[38], p.161), its expectation is

$$\mathrm{E}[z] := \int_\Omega z(\omega)\mathrm{d}\omega$$

with $\omega \in \Omega$. Note that the expectation of a $Z$-valued random function is a function in $Z$.

Let $Z_1, Z_2$ be two Hilbert space with inner products $(\cdot, \cdot)_{Z_1}$ and $(\cdot, \cdot)_{Z_2}$. With $v \in Z_1$ and $w \in Z_2$ we define the "$\circ$" mapping $\cdot \circ \cdot : Z_1 \times Z_2 \to \mathcal{L}(Z_2, Z_1)$ with $\mathcal{L}$ the set of linear bounded operators as

$$(v \circ w)w_1 := v\,(w, w_1)_{Z_2} \quad \forall w_1 \in Z_2.$$

The boundedness is a consequence of the Cauchy–Schwarz inequality. In the following we assume that all random functions are square integrable with respect to $\Omega$. We define the covariance operator $\mathrm{cov}(z_1, z_2) \in \mathcal{L}(Z_2, Z_1)$ of two $Z_1$- and $Z_2$-valued random functions $z_1 : \Omega \to Z_1$, $z_2 : \Omega \to Z_2$ as

$$\mathrm{cov}(z_1, z_2) := \mathrm{E}\left[ (z_1 - \mathrm{E}\,[z_1]) \circ (z_2 - \mathrm{E}\,[z_2]) \right].$$

We call $\mathrm{cov}(z) := \mathrm{cov}(z, z)$ the autocovariance operator of a $Z$-valued random function $z$. The autocovariance of a random function is a positive semidefinite self adjoint linear operator ($cf.$ [17]). We define the space of $Z$-valued square integrable random functions as $\mathcal{Z} := \mathcal{L}_{\in}(\otimes, \mathrm{Pr}; \mathcal{Z})$.

Let $s \in Z'$ be a linear functional in the dual space of $Z$ and $s_r \in Z$ its Riesz representation ($cf.$ [57], p. 90$f\!f\!.$) so that $s(v) = (s_r, v)$ for all $v \in Z$. We define the variance of $s$ applied to a random function $z \in \mathcal{Z}$ as

$$\mathrm{Var}\,(s(z)) := \mathrm{E}\left[ (s(z) - \mathrm{E}\,[s(z)])^2 \right].$$

Let $\mathrm{cov}(z)$ be the autocovariance operator of $z$, then one can easily prove ($cf.$ [17]) that

$$\mathrm{Var}\,(s(z)) = (s_r, \mathrm{cov}(z)s_r). \tag{2.1}$$

Later in the error analysis, the definition of the trace of an operator will be a useful tool. Hence, let $\{\psi_i\}_{i=1}^H$, $H \in \mathbb{N} \cup \{\infty\}$ be an orthonormal basis of the $H$-dimensional Hilbert space $Z_1 \subseteq Z$ inheriting the scalar product from $Z$ and be $A \in \mathcal{L}(Z)$. Then the trace of the operator $A$ with respect to $Z_1$ is defined as

$$\mathrm{trace}_{Z_1}\,(A) := \sum_{i=1}^H (\psi_i, A\psi_i). \tag{2.2}$$

The trace of an operator is invariant with the choice of the basis ($cf.$ [17]). The trace in Euclidian space is abbreviated $\mathrm{tr}\,(\boldsymbol{A}) := \mathrm{trace}_{\mathbb{R}^n}\,(\boldsymbol{A})$ for matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $n \in \mathbb{N}$.

The solution to the stochastically disturbed problem will be a time sequence of random functions. Therefore, we define the notion of a stochastic process. A $Z$-valued stochastic process is a map $\eta(\cdot, \cdot) : \mathbb{T} \times \Omega \to Z$ which is measurable on $\mathbb{T} \times \Omega$ with a uniform measure on $\mathbb{T}$ and the $\mathrm{Pr}$-measure on $\Omega$. For the elements of the stochastic process we use the abbreviation $\eta(t^k, \omega) =: \eta^k(\omega)$. Further, we call $\eta$ a white noise $Z$-valued random process if

$$\mathrm{E}\left[\eta^k(\omega)\right] = 0 \quad \text{and} \quad \mathrm{cov}(\eta^{k_1}(\omega), \eta^{k_2}(\omega)) = \delta_{k_1, k_2} R_{\eta\eta} \tag{2.3}$$

for all $k, k_1, k_2 = 0, \ldots, K$, with $R_{\eta\eta}$ a positive semidefinite linear self adjoint operator and $\delta$ the Kronecker symbol so that $\delta_{i,j} = 1$ for $i = j$ and $\delta_{i,j} = 0$ else. If additionally the $\eta^k(\omega)$ are normally distributed, we call $\eta$ a Gaussian white noise $Z$-valued random process.

Note, that $R$ usually symbolizes the correlation operator. However, the letter $R$ will be referred to as "covariance operator" in the case of white noise, as for white noise the covariance and the correlation operator coincide.

## 2.1. The stochastically forced parametrized parabolic PDE

Assuming a given parameter $\boldsymbol{\mu}$ stemming from a $p$-dimensional parameter domain $\mathcal{P} \subset \mathbb{R}^p$, we consider the parabolic parametrized time discrete partial differential equation with stochastic forcing in weak formulation

$$\begin{aligned}
m\left(z^k(\boldsymbol{\mu}, \omega), v\,; \boldsymbol{\mu}\right) + \Delta t\, a_0\left(z^k(\boldsymbol{\mu}, \omega), v\,; \boldsymbol{\mu}\right) = \\
m\left(z^{k-1}(\boldsymbol{\mu}, \omega), v\,; \boldsymbol{\mu}\right) + \Delta t\, b(v; \boldsymbol{\mu})u^k + \left(\eta^k(\omega), v\right)
\end{aligned} \tag{2.4}$$

for all $k \in \mathbb{K} := \{k \in \mathbb{N} | 1 \leq k \leq K\}$ with solution $z^k(\boldsymbol{\mu}, \omega) := z(t^k; \boldsymbol{\mu}, \omega) \in Z := \{v \mid v \in Z, v|_\Gamma = 0\}$, (random) initial condition $z^0(\boldsymbol{\mu}, \omega)$ and test functions $v \in Z$. Here, $m(\cdot, \cdot\,; \boldsymbol{\mu})$ and $a_0(\cdot, \cdot\,; \boldsymbol{\mu})$ are parametrized symmetric continuous bilinear forms and $a_0(\cdot, \cdot\,; \boldsymbol{\mu})$ is additionally coercive with coercivity constant $\alpha_0(\boldsymbol{\mu}) := \inf_{v \in Z} a_0(v, v\,; \boldsymbol{\mu}) / \|v\|^2$ with lower bound $\bar{\alpha}_0 \leq \alpha_0(\boldsymbol{\mu})$ and $b(\cdot; \boldsymbol{\mu})$ is a parametrized continuous linear form. The scalar input at time step $t^k$ is denoted $u^k := u(t^k) \in \mathbb{R}$ and $\eta^k$ are elements of a Gaussian white noise $Z$-valued random process $\eta$ with auto-covariance $\mathrm{cov}(\eta^k) = R_{\eta\eta}$. The bilinear forms $a_0$ and $m$ as well as the linear form $b$ are assumed to be parameter affine, hence, they can be represented by linear combinations

$$a_0(v, w\,; \boldsymbol{\mu}) = \sum_{q=1}^{Q_a} \Theta_a^q(\boldsymbol{\mu}) a_0^q(v, w), \qquad m(v, w\,; \boldsymbol{\mu}) = \sum_{q=1}^{Q_m} \Theta_m^q(\boldsymbol{\mu}) m^q(v, w), \tag{2.5}$$

$$b(v; \boldsymbol{\mu}) = \sum_{i=1}^{Q_b} \Theta_b^q(\boldsymbol{\mu}) b^q(v) \tag{2.6}$$

of parameter dependent coefficients $\Theta_a^q(\boldsymbol{\mu}), \Theta_m^q(\boldsymbol{\mu}), \Theta_b^q(\boldsymbol{\mu}) : \mathcal{P} \to \mathbb{R}$ and parameter independent bilinear and linear forms $a_0^q(v, w), m^q(v, w), b^q(v)$. In the remainder of this work we will often omit the realization $\omega$ and the parameter dependence $\boldsymbol{\mu}$ of $z$ in the notation and write $z^k = z^k(\boldsymbol{\mu}, \omega)$. The context will determine, if we note random variables or realization of random variables.

We assume that we can obtain and collect data from the system *via* measurements. Hence, we assume the existence of a (noise corrupted) scalar "measurement" output

$$y^k = s(z^k) + \vartheta^k \tag{2.7}$$

for all $k \in \mathbb{K}$ where $s \in Z'$ is a continuous linear output functional and $\vartheta^k$ are elements of the Gaussian white noise $\mathbb{R}$-valued stochastic process $\vartheta$ with covariance $R_{\vartheta\vartheta}$. The two stochastic processes $\eta$ and $\vartheta$ are considered to be independent. We denote with $s_r \in Z$ the Riesz representation of $s$.

In contrast to the available measurement data, we assume that we are interested in an "output of interest"

$$y_{int}^k = o(z^k)$$

with $o \in Z'$ and Riesz representation $o_r \in Z$. This output of interest can not be measured directly.

Note, that the extension to multiple inputs and outputs poses no conceptual difficulties for the following work. Yet, for the sake of readability we consider single input and single output systems.

## 2.2. The optimal filter in weak formulation

As the time-discrete infinite dimensional Kalman filter in weak formulation is a crucial ingredient in our theory and the foundation for the model reduction of the filter, we will outline shortly the derivation of the filter in this particular form. Also, in literature the Kalman filter for PDEs is often derived *via* infinite dimensional dynamical systems and semi groups (*e.g.* [24]). Here, we propose a different approach, which we consider to be more natural and better suited to the time-discrete and weak formulation of the PDE.

### 2.2.1. The filtering problem

The parametrized PDE in (2.4) could describe a real world engineering problem of a process being influenced by random disturbances $\eta^k$. Hence, the actual state of the system is unknown. In particular there is no way to evaluate the output of interest, as we assume it to be not directly measurable. The only information available about the system is the measurement data $y^k$ which is also subject to noise $\vartheta^k$. The overall objective is to find optimal estimates $\hat{z}^k$ for the noisy state $z^k$ of the system and optimal estimates $\hat{y}_{int}^k := o(\hat{z}^k)$ for the output of interest $o(z^k)$ at every time step $t^k$ using the noisy measurements $y^k$. Here, optimality is understood in the sense that the variance, or more specific the autocovariance $\mathrm{cov}(\tilde{z}^k, \tilde{z}^k)$ of the estimation error $\tilde{z}^k := z^k - \hat{z}^k$, is minimized. Hence, we are interested in minimizing the uncertainty in the state estimation. This requirement is formulated in the filtering problem.

**Problem 2.1** (The filtering problem)**.** Given the outputs $y^j$ for $j = 1, \ldots, k$ from (2.7), find an estimate $\hat{z}^k$ for the solution $z^k$ of (2.4) of the form

$$\hat{z}^k = \sum_{j=1}^{k} \mathcal{A}^j y^j, \ \mathcal{A}^j \in Z \tag{2.8}$$

with the property that

$$\left( \mathrm{E} \left[ \tilde{z}^k \circ \tilde{z}^k \right] v, v \right) \tag{2.9}$$

is minimized for all $v \in Z$.

Note, that (2.9) can be reformulated to: $\mathrm{E} \left[ v^*(\tilde{z}^k)^2 \right]$ *is minimized for all* $v^* \in Z'$. By this formulation it gets obvious, that the optimal filter provides output estimates $v^*(\hat{z}^k)$ with a minimal expected squared estimation error for every linear functional. Note also, that this is the filtering problem for linear time-discrete filters in Hilbert spaces. The equivalent filtering problem for linear time continuous filters in Hilbert spaces can be found in ([21], p. 124). The time discrete filtering problem, which is stated above, can be derived by discretizing the time continuous formulation in [21] in time.

The following Lemma will provide us with the necessary and sufficient condition to solve the filtering problem. In this Lemma we recognize the orthogonal projection Lemma, where the best approximation $\hat{z}^k$ is obtained by an orthogonal projection of $z^k$ onto an approximation space.

**Lemma 2.2.** *The function* $\hat{z}^k \in \mathcal{Z}$ *is a solution to the filtering problem if and only if*

$$\mathrm{E} \left[ \left( z^k - \hat{z}^k \right) y^j \right] = 0 \tag{2.10}$$

$\forall j = 1, \ldots, k$ *where* $z^k$ *is the solution to* (2.4) *and* $y^j \in \mathbb{R}$ *is the output* (2.7)*.*

*Proof.* The proof can be found in Appendix A.1.                                                                                □

**Remark 2.3.** We would like to comment on the equivalence to the orthogonal projection lemma. As shown in ([38], p. 163) the space of random functions with finite second moments together with the inner product $(\cdot, \cdot)_{\mathrm{Pr}} = \mathrm{E} \left[ \cdot \cdot \right]$ form a Hilbert space. Hence, the space $Y^k$ spanned by the observations $\{y^1, \ldots, y^k\}$ with $(\cdot, \cdot)_{\mathrm{Pr}}$ is a $k$-dimensional Hilbert space. While the random functions $z^k$ stem from the infinite dimensional Hilbert space of random functions $\mathcal{Z}$. Interpreting the Lemma above, the estimation $\hat{z}^k$ of the optimal filter (the Kalman filter) can be seen as the orthogonal projection of the solution $z^k$ from the infinite dimensional Space $\mathcal{Z}$ into the approximation space $Y^k$.

*2.2.2. Derivation of the Kalman filter in weak formulation*

For better readability of the following text, we introduce a new bilinear form

$$a \left( v, w; \boldsymbol{\mu} \right) := m \left( v, w \, ; \boldsymbol{\mu} \right) + \Delta t \, a_0 \left( v, w \, ; \boldsymbol{\mu} \right)$$

being coercive. Then, there exist unique operators $A, M \in \mathcal{L}(Z)$ defined by

$$(A(\boldsymbol{\mu})v, w) := a \left( v, w; \boldsymbol{\mu} \right) \tag{2.11}$$

$$(M(\boldsymbol{\mu})v, w) := m \left( v, w \, ; \boldsymbol{\mu} \right) \ \forall v, w \in Z. \tag{2.12}$$

(This is a consequence of the Riesz representation theorem *cf*. [57], p. 90*ff*.). In the following we will often omit the parameter dependence of $A = A(\boldsymbol{\mu})$ and $M = M(\boldsymbol{\mu})$. With $a$ being a coercive and continuous bilinear form, we know by the Lax–Milgram Theorem that there is a unique inverse operator $A^{-1} \in \mathcal{L}(Z)$ with $A^{-1}A = I$ with $I$ the identity. The operator $A$ is self adjoint as the bilinear form $a \left( \cdot, \cdot; \boldsymbol{\mu} \right)$ is symmetric: $(Av, w) = a \left( v, w; \boldsymbol{\mu} \right) = a \left( w, v; \boldsymbol{\mu} \right) = (Aw, v) = (v, Aw)$. As the solution $z^k$ as well as the estimation $\hat{z}^k$ are random functions for all $k = 0, \ldots, K$, the estimation error $\tilde{z}^k = z^k - \hat{z}^k$ is also a $Z$-valued random function.

As the noise $\eta$ is assumed to be a Gaussian white noise random process and in the evolution equation only linear operations are applied, the estimation error $\tilde{z}^k$ will also be normally distributed. Furthermore, we will see later, that in the case of an optimal filter the expectation of the estimation error will be zero. We define the estimation error covariance operator as

$$P^k := \text{cov}\left(\tilde{z}^k, \tilde{z}^k\right). \tag{2.13}$$

Using Lemma 2.2 we can derive the time-discrete Kalman filter for parabolic partial differential equations in weak formulation (omitting the parameter dependence $\boldsymbol{\mu}$ in auxilliary quantities):

**Proposition 2.4** (Optimal filter). *The solution $\hat{z}^k$ to the time discrete PDE in weak formulation*

$$m\left(\hat{z}^k, v\,;\boldsymbol{\mu}\right) + \Delta t\, a_0\left(\hat{z}^k, v\,;\boldsymbol{\mu}\right) = m\left(\hat{z}^{k-1}, v\,;\boldsymbol{\mu}\right) + \Delta t\, b(v;\boldsymbol{\mu})u^k + l^k(v)\left(y^k - s(\hat{z}_p^k)\right) \tag{2.14}$$

*for all $v \in Z$ and for all $k = 1, \ldots, k$ with an initial estimate $\hat{z}^0 = \text{E}\left[z^0\right]$ is the unique solution to the filtering problem. Here, the propagated estimate $\hat{z}_p^k$ is the solution to*

$$m\left(\hat{z}_p^k, v\,;\boldsymbol{\mu}\right) + \Delta t\, a_0\left(\hat{z}_p^k, v\,;\boldsymbol{\mu}\right) = m\left(\hat{z}^{k-1}, v\,;\boldsymbol{\mu}\right) + \Delta t\, b(v;\boldsymbol{\mu})u^k \qquad \forall v \in Z \tag{2.15}$$

*with $\hat{z}_p^0 = \hat{z}^0$ and the Kalman gain $l^k(v)$ is a linear form defined by*

$$l^k(v) := \left(v, P_p^k s_r\right) \cdot \left(\left(A^{-1}s_r, P_p^k s_r\right) + R_{\vartheta\vartheta}\right)^{-1} \qquad \forall v \in Z \tag{2.16}$$

*with the propagated error covariance*

$$\left(w, P_p^k v\right) := \left(Mw, P^{k-1}MA^{-1}v\right) + \left(w, R_{\eta\eta}A^{-1}v\right) \qquad \forall v, w \in Z \tag{2.17}$$

*and the evolution of the error covariance*

$$\left(w, P^k v\right) = \left(A^{-1}w, P_p^k v\right) - l^k(A^{-1}w)\left(A^{-1}s_r, P_p^k v\right) \qquad \forall v, w \in Z \tag{2.18}$$

*with an initial error covariance $P^0 = \text{E}\left[\tilde{z}^0 \circ \tilde{z}^0\right]$ and the Riesz representation $s_r \in Z$ of the output functional.*

*Proof.* The proof can be found in the Appendix A.2. $\qquad\square$

## 3. REDUCED BASIS KALMAN FILTER

After stating the Kalman filter in weak formulation in infinite dimensional Hilbert spaces, we will now handle the issue of numerical approximation in finite dimensional Hilbert spaces and present a rapidly computable approximation of the optimal filter in algebraic form. We introduce a finite dimensional Hilbert space $Z_h \subset Z$ of very high dimension $H = \dim(Z_h)$, where $Z_h$ inherits the scalar product and norm from $Z$. Performing a Galerkin projection of (2.4) onto $Z_h$ we obtain

$$m\left(z_h^k(\boldsymbol{\mu},\omega), v\,;\boldsymbol{\mu}\right) + \Delta t\, a_0\left(z_h^k(\boldsymbol{\mu},\omega), v\,;\boldsymbol{\mu}\right) = \tag{3.1}$$
$$m\left(z_h^{k-1}(\boldsymbol{\mu},\omega), v\,;\boldsymbol{\mu}\right) + \Delta t\, b(v;\boldsymbol{\mu})u^k + \left(\eta^k(\omega), v\right)$$

for all $v \in Z_h$ and for $k = 1, \ldots, K$. The solution $z_h^k$ could be a finite element solution for example, and we assume this "true" solution to be sufficiently accurate, such that the finite element discretization error is negligible compared to our reduction error in the following. The measurement output of the "true" system is modeled by

$$y_h^k := s(z_h^k) + \vartheta^k \tag{3.2}$$

and the output of interest is simply $y_{h,int}^k := o(z_h^k)$.

The optimal filter for the high dimensional stochastically forced system (3.1), (3.2) can be obtained by projecting the infinite dimensional Kalman filter (2.14)–(2.18) onto the finite dimensional space $Z_h$. However, then we would obtain very high dimensional matrix equations. Performing state estimations using this high dimensional Kalman filter could be computationally very expensive. In order to allow rapid state estimations, we will instead use a reduced basis approximation of the Kalman filter. For that purpose, let $Z_N = \text{span}\{\varphi_1, \ldots, \varphi_N\} \subseteq Z_h$ be the reduced basis approximation space of dimension $\dim(Z_N) = N \ll \dim(Z_h)$ spanned by the orthonormal basis $\{\varphi_1, \ldots, \varphi_N\}$ with $\varphi_i \in Z_h$ and $(\varphi_i, \varphi_j) = \delta_{i,j}$ for all $i, j = 1, \ldots, N$. The ansatz for the RBKF estimation is $\hat{z}_N^k = \sum_{i=1}^N \hat{z}_{N,i}^k \varphi_i$ with coefficients $\hat{z}_{N,i}^k \in \mathbb{R}$. We define the coefficient vector $\hat{\boldsymbol{z}}_N^k := (\hat{z}_{N,1}^k, \ldots, \hat{z}_{N,N}^k)^T \in \mathbb{R}^N$. A Galerkin projection of the infinite dimensional Kalman filter from (2.14)–(2.18) on $Z_N$ results in the algebraic RBKF formulation: we define the projected operator matrices and projected linear forms by

$$
\begin{aligned}
(\boldsymbol{A}_N(\boldsymbol{\mu}))_{i,j} &:= a(\varphi_j, \varphi_i; \boldsymbol{\mu}), \quad (\boldsymbol{M}_N(\boldsymbol{\mu}))_{i,j} := m(\varphi_j, \varphi_i; \boldsymbol{\mu}) \\
(\boldsymbol{B}_N(\boldsymbol{\mu}))_i &:= \Delta t\, b(\varphi_i; \boldsymbol{\mu}), \quad (\boldsymbol{C}_N)_i := s(\varphi_i) \\
(\boldsymbol{R}_{\eta\eta,N})_{i,j} &:= (\varphi_i, R_{\eta\eta}\varphi_j), \quad (\boldsymbol{P}_N^0)_{i,j} := \left(\varphi_i, \mathrm{E}\left[\tilde{z}^0 \circ \tilde{z}^0\right]\varphi_j\right)
\end{aligned}
\tag{3.3}
$$

for all $i, j = 1, \ldots, N$ and obtain the RBKF state estimations by solving

$$
\boldsymbol{A}_N \hat{\boldsymbol{z}}_N^k = \boldsymbol{M}_N \hat{\boldsymbol{z}}_N^{k-1} + \boldsymbol{B}_N^k u^k + \boldsymbol{L}_N^k \left(y_h^k - \boldsymbol{C}_N^T \hat{\boldsymbol{z}}_{N,p}^k\right)
\tag{3.4}
$$

for all $k \in \mathbb{K}$ with initial condition $\hat{z}_N^0$ satisfying $\left(\hat{z}_N^0, v_N\right) = \left(\mathrm{E}\left[z^0\right], v_N\right) \forall v_N \in Z_N$ and the propagation

$$
\boldsymbol{A}_N \hat{\boldsymbol{z}}_{N,p}^k = \boldsymbol{M}_N \hat{\boldsymbol{z}}_N^{k-1} + \boldsymbol{B}_N^k u^k
\tag{3.5}
$$

with $\hat{\boldsymbol{z}}_{N,p}^0 = \hat{\boldsymbol{z}}_N^0$. The RBKF gain vector can be calculated *via*

$$
\boldsymbol{L}_N^k = \boldsymbol{P}_{N,p}^k \boldsymbol{C}_N \left(\boldsymbol{C}_N^T \boldsymbol{A}_N^{-T} \boldsymbol{P}_{N,p}^k \boldsymbol{C}_N + R_{\vartheta\vartheta}\right)^{-1}
\tag{3.6}
$$

with

$$
\boldsymbol{P}_{N,p}^k := \boldsymbol{M}_N \boldsymbol{P}_N^{k-1} \boldsymbol{M}_N \boldsymbol{A}_N^{-1} + \boldsymbol{R}_{\eta\eta,N} \boldsymbol{A}_N^{-1}
\tag{3.7}
$$

and the covariance matrix of the estimation error

$$
\boldsymbol{P}_N^k = \boldsymbol{A}_N^{-T} \boldsymbol{P}_{N,p}^k - \boldsymbol{A}_N^{-T} \boldsymbol{L}_N^k \boldsymbol{C}_N^T \boldsymbol{A}_N^{-T} \boldsymbol{P}_{N,p}^k.
\tag{3.8}
$$

As all involved vectors and matrices are only of low dimension $N$, the whole reduced order Kalman filter scheme can be calculated considerably faster than in the high dimensional case. It is of importance to point out, that the measurement output $y_h^k$ from the true system (3.2) is used in the RBKF formulation (3.4).

**Remark 3.1.** It can be shown, that the projection of the infinite dimensional optimal filter onto a finite dimensional space $Z_N$ results in an optimal filter for the finite dimensional system resulting from a projection of (2.4) onto $Z_N$. For proving this statement it suffices to show that the finite dimensional filter obtained by projection satisfies Lemma 2.2.

### 3.1. Basis generation

The wise choice of the reduced basis approximation space $Z_N$ is crucial in model reduction using the reduced basis method, as it determines the overall quality of the approximation procedure. The space should be chosen optimally regarding two main criteria: (I) the reduced space is supposed to be of very low dimension in order to allow rapid solves; and (II) the reduced space should provide the necessary degrees of freedom to find accurate solutions with low approximation errors.

One of the most common methods in reduced basis methods is to create a reduced approximation space using a Greedy algorithm [55], or in time dependent problems the POD-Greedy algorithm [28]. Convergence rates of these basis construction methods have been shown in [3, 7, 26]. There exist also various adaptive extension techniques as elaborated in [18, 20, 27].

In our case of a stochastically forced problem, the solution is deformed and disturbed by additive white noise. It is crucial to take this noise effect into account during the basis generation process. In the experiments in Section 5.1 we will demonstrate numerically the advantage for the following basis generation technique for the model reduction of systems corrupted by white noise.

A crucial ingredient to the construction of the reduced basis is the Karhunen–Loève (KL) expansion to reformulate the noise term. For details on the KL expansion we refer to [56]. We assume that $\xi_i \in Z$ with $i = 1, \ldots, d$, $d \in \mathbb{N} \cup \{\infty\}$, are the eigenfunctions of the white noise covariance operator $R_{\eta\eta}$, so that

$$R_{\eta\eta}\xi_i = \lambda_i\xi_i$$

for eigenvalues $\lambda_i \in \mathbb{R}$. Then the white noise random function $\eta^k$ can be reformulated as

$$\eta^k = \sum_{i=1}^d \sqrt{\lambda_i}\xi_i\nu_i^k \tag{3.9}$$

with $\nu_i^k \backsim N(0,1)$ a normally distributed $\mathbb{R}$-valued random variable ($cf.$ [56]). We assume, that the eigenfunctions and eigenvalues are hierarchically ordered, so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. In the following we will choose a number $n_{KL} \in \mathbb{N}$ where the KL-expansion is cut off, so that we approximate $\eta^k \approx \sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\xi_i\nu_i^k$. Note, that the number of KL-modes used for the approximation of the noise characteristics has to be chosen carefully.

The idea behind our basis generation approach is the following: if we plug in the KL-expansion from (3.9) into the stochastically forced PDE from (2.4), it gets obvious that we actually have a linear time invariant (LTI) system with deterministic input $u^k$ and random inputs $\nu_i^k$. Hence, the "noisy" part $\eta^k \approx \sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\xi_i\nu_i^k$ can be interpreted as a (random) input to our linear system. Duhamel's principle states, that the solution of an LTI system to any input can be found by a convolution of the impulse response of the system with the given input. Consequently, if we design the reduced basis space in a way so that it approximates well the impulse solutions, then it will be possible to represent well solutions for any input – deterministic and random.

To obtain those deterministic and random impulse solutions for the basis generation, we first define the "detailed" impulse solution $z_{h,imp}^k(\boldsymbol{\mu}, n) \in Z_h$ with impulse number $n = 1, \ldots, n_{KL} + 1$ as solution to

$$a\left(z_{h,imp}^k(\boldsymbol{\mu}, n), v; \boldsymbol{\mu}\right) = m\left(z_{h,imp}^{k-1}(\boldsymbol{\mu}, n), v; \boldsymbol{\mu}\right) + \Delta t\, b(v; \boldsymbol{\mu})u_{imp}^k(n) + \sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\,(v, \xi_i)\,\nu_{i,imp}^k(n) \tag{3.10}$$

for all $v \in Z_h$ with initial condition $z_{h,imp}^0(\boldsymbol{\mu}, n) = 0$ and impulse inputs $u_{imp}^k(n) := \delta_{k,1} \cdot \delta_{n,1}$, $\nu_{i,imp}^k(n) := \delta_{k,1} \cdot \delta_{n-1,i}$. While the reduced impulse solution $z_{N,imp}^k(\boldsymbol{\mu}, n) \in Z_N$ is the Galerkin projection of (3.10) onto $Z_N$. Note, that in the RB approximation of the impulse response, we use the same number of KL-modes as in the detailed impulse response. Next, we assume the existence of a rapidly computable error bound $\Delta(\boldsymbol{\mu}; n) \geq \max_{k \in \mathbb{K}} \|z_{h,imp}^k(\boldsymbol{\mu}, n) - z_{N,imp}^k(\boldsymbol{\mu}, n)\|$. There exist various approaches to derive these kind of bounds ($e.g.$ [25, 28]). Also, we will present such an error bound later in Section 4.3.

The "Noise-POD-Greedy" basis generation algorithm for stochastically forced models works as follows: we start with an initial set of reduced basis functions $\Phi = \Phi_0$. We also choose a training set of parameters $\mathcal{M}_{train} \subset \mathcal{P}$ and the number of KL-modes $n_{KL}$ we wish to consider for basis generation. To control the quality of the reduced basis space, we fix an approximation tolerance $\varepsilon_{tol} > 0$ and a maximal basis size $N_{\max} \in \mathbb{N}$. We start the procedure by calculating the error estimate $\Delta(\boldsymbol{\mu}, n)$ for all impulse solutions for all $\boldsymbol{\mu} \in \mathcal{P}$ and $n = 1, \ldots, n_{KL} + 1$ (impulses on the deterministic input $u$ and the first $n_{KL}$ modes). It is important to mention, that for computing $\Delta(\boldsymbol{\mu}, n)$ only low dimensional calculations are required in order to enable a

quick search over all parameters and inputs. Here, we choose the pair of parameter $\boldsymbol{\mu}^*$ and $n^*$ producing the highest error estimate $\Delta(\boldsymbol{\mu}^*, n^*)$. For this pair of parameter and impulse number, we calculate the detailed impulse solution $z_{h,imp}(\boldsymbol{\mu}^*, n^*)$. *Via* a projection $\Pi_\Phi(z_{h,imp}(\boldsymbol{\mu}^*, n^*))$ onto span$\{\Phi\}$ and subsequently building the orthogonal complement $z_\perp^k = z_{h,imp}(\boldsymbol{\mu}^*, n^*) - \Pi_\Phi(z_{h,imp}(\boldsymbol{\mu}^*, n^*))$ we pull out the novelty that this impulse solution trajectory can bring to the existing reduced basis. To compress this new information in an optimal way, we perform a POD (*cf.* [35], Sect. 2) over this orthogonal complement of the solution trajectory and take the first POD mode to enrich the reduced basis. Now that we have added a new basis function to the reduced basis $\Phi$, we restart searching for the pair $(\boldsymbol{\mu}^*, n^*)$ maximizing $\Delta(\boldsymbol{\mu}, n)$ using the new enhanced reduced basis $\Phi$ and we verify if the quality constraints are fulfilled, *i.e.* if $\Delta(\boldsymbol{\mu}^*, n^*) \leq \varepsilon_{tol}$. We continue until one of the constraints is satisfied and we ultimately obtain a reduced basis space $Z_N = \text{span}\{\Phi\}$ suited to approximate well the solutions of the stochastically forced partial differential equation. The basis generation algorithm is summarized in Algorithm 1.

**Algorithm 1.** The Noise-POD-Greedy procedure for building up reduced basis spaces in case of stochastically forced problems.

---
NOISE-POD-GREEDY$(\Phi_0, \mathcal{M}_{train}, \varepsilon_{tol}, N_{\max}, n_{KL})$
1   $\Phi := \Phi_0$
2   **repeat**
3        $(\boldsymbol{\mu}^*, n^*) := \arg\max_{\boldsymbol{\mu} \in \mathcal{M}_{train}, n=1,\ldots,n_{KL}+1} \Delta(\boldsymbol{\mu}, n)$
4        **if** $\Delta(\boldsymbol{\mu}^*, n^*) > \varepsilon_{tol}$
5          **then**
6            $z_{h,imp}(\boldsymbol{\mu}^*, n^*) := \text{CALCULATEDETAILEDIMPULSESOLUTION}(\boldsymbol{\mu}^*, n^*)$
7            $\Phi = \Phi \cup \text{POD}(z_{h,imp}(\boldsymbol{\mu}^*, n^*) - \Pi_\Phi z_{h,imp}(\boldsymbol{\mu}^*, n^*), 1)$
8   **until** $\Delta(\boldsymbol{\mu}^*, n^*) \leq \varepsilon_{tol}$ **or** $|\Phi| \geq N_{\max}$
9   **return** $\Phi$

---

## 3.2. Computational procedure and offline-online decomposition

The computational procedure for the RBKF is divided into a preparing offline phase and an online phase, where the calculations for the state estimations are performed. The objective of this separation is to have the actual calculations for the RBKF independent of the size of the high dimensional space $Z_h$ and only depending on the number of dimensions of the low dimensional space $Z_N$. Thereby, rapid state estimations – or even real-time state estimations – can be assured.

In the offline phase the reduced basis is constructed as described in Section 3.1. Furthermore, the parameter independent components of the bilinear and linear forms are projected onto the reduced space $Z_N$ by computing $(\boldsymbol{A}_{0,N}^{q_a})_{i,j} := a_0^{q_a}(\varphi_j, \varphi_i)$, $(\boldsymbol{M}_N^{q_m})_{i,j} := m^{q_m}(\varphi_j, \varphi_i)$, $(\boldsymbol{B}_N^{q_b})_i := \Delta t \, b^{q_b}(\varphi_i)$ for all $i, j = 1 \ldots, N$ and $q_a = 1, \ldots, Q_a$, $q_m = 1, \ldots, Q_m$ and $q_b = 1, \ldots, Q_b$.

Then, in the online phase, the required matrices $\boldsymbol{A}_N(\boldsymbol{\mu})$, $\boldsymbol{M}_N(\boldsymbol{\mu})$ and the vector $\boldsymbol{B}_N(\boldsymbol{\mu})$ can be assembled quickly for a given parameter $\boldsymbol{\mu}$ by taking advantage of the parameter affinity of the bilinear and linear forms *via*

$$\boldsymbol{A}_N(\boldsymbol{\mu}) = \sum_{q=1}^{Q_m} \Theta_m^q(\boldsymbol{\mu}) \boldsymbol{M}_N^q + \Delta t \sum_{q=1}^{Q_a} \Theta_a^q(\boldsymbol{\mu}) \boldsymbol{A}_{0,N}^q, \qquad (3.11)$$

$$\boldsymbol{M}_N(\boldsymbol{\mu}) = \sum_{q=1}^{Q_m} \Theta_m^q(\boldsymbol{\mu}) \boldsymbol{M}_N^q, \quad \boldsymbol{B}_N(\boldsymbol{\mu}) = \sum_{q=1}^{Q_b} \Theta_b^q(\boldsymbol{\mu}) \boldsymbol{B}_N^q. \qquad (3.12)$$

After having assembled the parameter dependent quantities, RBKF estimations can be performed. Note, that this technique even allows to treat the more general situation of time-varying parameters. The assembling (3.11), (3.12) is then performed at every time step.

Note also, that the offline phase can be computationally expensive. On the other hand, after having completed the preparations, the computations in the online phase are independent of high dimensional quantities and provide rapid state estimations even for varying parameters.

## 4. Error estimation

Let us first discuss, why it is necessary and useful to handle the issue of error estimation for the RBKF. When re-examining the filtering problem we see, that it is formulated for a specific combination of a system and measurement data from the system. It becomes obvious that an "optimal filter" is only optimal considering this combination of system and data. Hence, the time discrete Kalman filter in weak formulation in Proposition 2.4 is an optimal filter for the semi-discrete system (2.4) with outputs $y^k$ from (2.7). As pointed out in Remark 3.1, also the projection of the infinite dimensional Kalman filter onto the space $Z_h$ would be an optimal filter for the finite dimensional system (3.1) with outputs $y_h^k$. However, the reduced basis Kalman filter being steered by the measurement outputs $y_h^k$ from the true system (3.1) does not give optimal state estimations for the true noisy system. It only provides an approximation of an optimal filter for that system.

The fact that the RBKF is not an optimal filter has an impact on the evolution of the estimation error. While in the optimal filter case, the mean of the estimation error is known to be zero and the error covariance operator $P^k$ is explicitly provided by the Kalman filter scheme, we can not assume the mean estimation error $\mathrm{E}\left[z_h^k - \hat{z}_N^k\right]$ of the RBKF to be zero and the estimation error covariance of the RBKF $\mathrm{cov}(z_h^k - \hat{z}_N^k)$ to be given by $\boldsymbol{P}_N^k$ in (3.8). This problem of having only an approximative Kalman filter can cause divergence in the state estimations and even lead to unstable filters ($cf$. [23, 46]).

At the end of this section we will show, how the stability of the RB Kalman filter can be verified *a priori*. But first, we consider a finite time horizon and derive an upper bound for the expectation of the estimation error of the output of interest $\mathrm{E}\left[o(z_h^k - \hat{z}_N^k)\right]$ and the variance of this estimation error $\mathrm{Var}\left(o(z_h^k - \hat{z}_N^k)\right)$. For computing these bounds only low dimensional calculations will be needed, hence, a certification of the RB Kalman filter estimations is possible in the online phase. This clearly is an advantage compared to other reduced order filters. Note, that we assumed the high dimensional "true" model to be accurate and the error between the analytic solution $z^k$ and the high dimensional approximation $z_h^k$ to be negligible. Hence, in the following we will only be interested in the error between the truth solution $z_h^k$ and the RBKF estimations $\hat{z}_N^k$.

As this section is rather long, we will give a short overview of its structure. In order to have a tight error bound we will apply primal dual error estimation techniques, which are based mainly on [25], where primal-dual error estimation was introduced for time-dependent RB problems. The dual problems as well as a virtual noisy RB model will be introduced in Section 4.1. Then, the estimation error will be decomposed into four parts in Section 4.2 and in the subsequent sections a bound for the expectation and the variance of all four error components will be derived. Then we will comment on possible simplifications for the error estimator. Finally, in Section 4.6 we will see, how the stability of the RBKF can be verified *a priori via* an observability criterion.

### 4.1. Preliminaries

For the error analysis of the RB Kalman filter we introduce the noisy reduced basis model

$$m\left(z_N^k(\boldsymbol{\mu},\omega), v_N\,;\boldsymbol{\mu}\right) + \Delta t\, a_0\left(z_N^k(\boldsymbol{\mu},\omega), v_N\,;\boldsymbol{\mu}\right) = \tag{4.1}$$
$$m\left(z_N^{k-1}(\boldsymbol{\mu},\omega), v_N\,;\boldsymbol{\mu}\right) + \Delta t\, b(v_N;\boldsymbol{\mu})u^k + \left(\eta^k(\omega), v_N\right)$$

for all $v_N \in Z_N$ and for all $k \in \mathbb{K}$ resulting from a Galerkin projection of (2.4) onto $Z_N$ and initial condition $z_N^0$ defined *via* $\left(z_N^0 - z_h^0, v_N\right) = 0$ for all $v_N \in Z_N$. We define the measurement output and the output of interest of the reduced system as

$$y_N^k := s(z_N^k) + \vartheta^k \quad \text{and} \quad y_{N,int}^k := o(z_N^k). \tag{4.2}$$

Note, that (4.1) can not be practically computed, as the noise is inaccessible. Therefore, we denote this system to be "virtual" and it will only be involved in the error analysis. The same holds for the residual defined as

$$\operatorname{Res}^k(v) := -a\left(z_N^k, v; \boldsymbol{\mu}\right) + m\left(z_N^{k-1}, v\ ; \boldsymbol{\mu}\right) + \Delta t\, b(v; \boldsymbol{\mu})u^k + \left(\eta^k, v\right). \tag{4.3}$$

Next, we introduce an optimal (and also virtual) RB Kalman filter estimation $\hat{z}_N^k(y_N) := \sum_{i=1}^N (\hat{\boldsymbol{z}}_N^k(y_N))_i \varphi_i$ with coefficient vector $\hat{\boldsymbol{z}}_N^k(y_N) \in \mathbb{R}^N$ obtained by solving

$$\boldsymbol{A}_N \hat{\boldsymbol{z}}_N^k(y_N) = \boldsymbol{M}_N \hat{\boldsymbol{z}}_N^{k-1}(y_N) + \boldsymbol{B}_N^k u^k + \bar{\boldsymbol{L}}_N^k \left(y_N^k - \boldsymbol{C}_N^T (\boldsymbol{A}_N^{-1} \boldsymbol{M}_N \hat{\boldsymbol{z}}_N^{k-1}(y_N) + \boldsymbol{A}_N^{-1} \boldsymbol{B}_N^k u^k)\right) \tag{4.4}$$

and

$$\bar{\boldsymbol{L}}_N^k \ = \ \bar{\boldsymbol{P}}_{N,p}^k \boldsymbol{C}_N \left(\boldsymbol{C}_N^T \boldsymbol{A}_N^{-T} \bar{\boldsymbol{P}}_{N,p}^k \boldsymbol{C}_N + R_{\vartheta\vartheta}\right)^{-1} \tag{4.5}$$

$$\bar{\boldsymbol{P}}_{N,p}^k := \boldsymbol{M}_N \bar{\boldsymbol{P}}_N^{k-1} \boldsymbol{M}_N \boldsymbol{A}_N^{-1} + \boldsymbol{R}_{\eta\eta,N} \boldsymbol{A}_N^{-1} \tag{4.6}$$

$$\bar{\boldsymbol{P}}_N^k \ = \ \boldsymbol{A}_N^{-T} \bar{\boldsymbol{P}}_{N,p}^k - \boldsymbol{A}_N^{-T} \bar{\boldsymbol{L}}_N^k \boldsymbol{C}_N^T \boldsymbol{A}_N^{-T} \bar{\boldsymbol{P}}_{N,p}^k \tag{4.7}$$

iteratively for all $k \in \mathbb{K}$, with initial estimation $(\hat{\boldsymbol{z}}_N^0(y_N))_i = \left(\mathrm{E}\left[z_h^0\right], \varphi_i\right)$ and $\bar{\boldsymbol{P}}_N^0 = \boldsymbol{P}_N^0$. As this reduced basis Kalman filter estimation is steered by the reduced outputs $y_N^k$, we denote $\hat{\boldsymbol{z}}_N^k(y_N)$ as a function of the reduced data $y_N$. Hence, we distinguish two different RB Kalman filter estimations. Firstly, the real RB Kalman filter estimation, now noted $\hat{z}_N^k(y_h)$ solution of (3.4)–(3.8) and steered by the true measurement output $y_h^k$ from (3.2). Secondly, the (virtual) optimal RB Kalman filter estimation $\hat{z}_N^k(y_N)$ steered by the (virtual) output $y_N^k$ from the noisy reduced basis model (4.2). Considering Remark 3.1, we know that the $\hat{z}_N^k(y_N)$ for all $k \in \mathbb{K}$ are optimal state estimations.

In order to obtain a sharp error bound for useful certification of the RBKF, we propose to apply a primal-dual approach inspired by [25]. Therefor we introduce (finite dimensional) dual problems for the measurement output and for the output of interest. The dual solutions $z_{\mathrm{du}}^k := z_{\mathrm{du}}(t^k; \boldsymbol{\mu}) \in Z_h$ and $z_{\mathrm{du},o}^k = z_{\mathrm{du},o}(t^k, \boldsymbol{\mu}) \in Z_h$ solve

$$m\left(z_{\mathrm{du}}^k, v\ ; \boldsymbol{\mu}\right) + \Delta t\, a_0\left(z_{\mathrm{du}}^k, v\ ; \boldsymbol{\mu}\right) = m\left(z_{\mathrm{du}}^{k+1}, v\ ; \boldsymbol{\mu}\right)$$

$$\text{and} \quad m\left(z_{\mathrm{du},o}^k, w\ ; \boldsymbol{\mu}\right) + \Delta t\, a_0\left(z_{\mathrm{du},o}^k, w\ ; \boldsymbol{\mu}\right) = m\left(z_{\mathrm{du},o}^{k+1}, w\ ; \boldsymbol{\mu}\right)$$

for time indices $k = K, \ldots, 1$ with end time conditions

$$m\left(z_{\mathrm{du}}^{K+1}, v\ ; \boldsymbol{\mu}\right) = s(v) \quad , \quad m\left(z_{\mathrm{du},o}^{K+1}, w\ ; \boldsymbol{\mu}\right) = o(w) \tag{4.8}$$

for all $v, w \in Z_h$. Note, that both dual solutions $z_{\mathrm{du}}^k$ and $z_{\mathrm{du},o}^k$ are non-random functions.

Next, reduced basis approximations of the dual solutions in the low dimensional spaces $Z_{N,\mathrm{du}} = \mathrm{span}\{\varphi_{\mathrm{du},1}, \ldots, \varphi_{\mathrm{du},N_{\mathrm{du}}}\}$ and $Z_{N,\mathrm{du},o} = \mathrm{span}\{\varphi_{\mathrm{du},o,1}, \ldots, \varphi_{\mathrm{du},o,N_{\mathrm{du},o}}\}$ of dimensions $\dim(Z_{N,\mathrm{du}}) = N_{\mathrm{du}}$ and $\dim(Z_{N,\mathrm{du},o}) = N_{\mathrm{du},o}$ are introduced (for building both reduced basis spaces, standard POD-Greedy procedures [28] can be performed). The reduced dual solutions $z_{N,\mathrm{du}}^k := z_{N,\mathrm{du}}(t^k; \boldsymbol{\mu}) \in Z_{N,\mathrm{du}}$ and $z_{N,\mathrm{du},o}^k := z_{N,\mathrm{du},o}(t^k; \boldsymbol{\mu}) \in Z_{N,\mathrm{du},o}$ solve the reduced dual problems

$$m\left(z_{N,\mathrm{du}}^k, v_N\ ; \boldsymbol{\mu}\right) + \Delta t\, a_0\left(z_{N,\mathrm{du}}^k, v_N\ ; \boldsymbol{\mu}\right) = m\left(z_{N,\mathrm{du}}^{k+1}, v_N\ ; \boldsymbol{\mu}\right),$$

$$m\left(z_{N,\mathrm{du},o}^k, w_N\ ; \boldsymbol{\mu}\right) + \Delta t\, a_0\left(z_{N,\mathrm{du},o}^k, w_N\ ; \boldsymbol{\mu}\right) = m\left(z_{N,\mathrm{du},o}^{k+1}, w_N\ ; \boldsymbol{\mu}\right)$$

for time indices $k = K, \ldots, 1$ with final solutions

$$m\left(z_{N,\mathrm{du}}^{K+1}, v_N\ ; \boldsymbol{\mu}\right) = s(v_N) \quad \text{and} \quad m\left(z_{N,\mathrm{du},o}^{K+1}, w_N\ ; \boldsymbol{\mu}\right) = o(w_N) \tag{4.9}$$

for all $v_N \in Z_{N,\mathrm{du}}$ and $w_N \in Z_{N,\mathrm{du},o}$. For the error analysis, the dual residuals

$$\mathrm{Res}^k_{\mathrm{du}}(v) := m\left(z^{k+1}_{N,\mathrm{du}} - z^k_{N,\mathrm{du}}, v \; ; \boldsymbol{\mu}\right) - \Delta t \, a_0 \left(z^k_{N,\mathrm{du}}, v \; ; \boldsymbol{\mu}\right) \tag{4.10}$$

and

$$\mathrm{Res}^k_{\mathrm{du},o}(w) := m\left(z^{k+1}_{N,\mathrm{du},o} - z^k_{N,\mathrm{du},o}, w \; ; \boldsymbol{\mu}\right) - \Delta t \, a_0 \left(z^k_{N,\mathrm{du},o}, w \; ; \boldsymbol{\mu}\right) \tag{4.11}$$

are defined for all $v \in Z_h$ and $w \in Z_h$.

The error will be measured in a spatio-temporal energy norm ($cf.$ [25]) defined as

$$|||(v^i)^k_{i=1}|||^2 = m\left(v^k, v^k \; ; \boldsymbol{\mu}\right) + \Delta t \sum_{i=1}^{k} a_0 \left(v^i, v^i \; ; \boldsymbol{\mu}\right) \tag{4.12}$$

for all $(v^i)^k_{i=1} \in (Z)^k$. Note, that instead of writing $|||(v^i)^k_{i=1}|||$ we may also write $|||v^k|||$ and mean thereby the energy norm over the trajectory up to $v^k$. Using this formulation, the appropriate norm for trajectories in the dual space is defined for all $(w)^K_{i=k} \in (Z)^{K-k+1}$ as:

$$|||w^k|||^2_{\mathrm{du}} = m\left(w^k, w^k \; ; \boldsymbol{\mu}\right) + \Delta t \sum_{i=k}^{K} a_0 \left(w^i, w^i \; ; \boldsymbol{\mu}\right). \tag{4.13}$$

## 4.2. Error decomposition

The key step in the error estimation of the RBKF is the decomposition of the estimation error. The basic idea for this decomposition is the reformulation of the estimation error as

$$z^k_h - \hat{z}^k_N(y_h) = \underbrace{z^k_h - z^k_N}_{(I)} + \underbrace{z^k_N - \hat{z}^k_N(y_N)}_{(II)} + \underbrace{\hat{z}^k_N(y_N) - \hat{z}^k_N(y_h)}_{(III)}. \tag{4.14}$$

Hence, the estimation error can be written as a sum of: (I) the error between the noisy true system and the (virtual) noisy reduced system, (II) the estimation error between the reduced noisy system and the reduced basis Kalman filter steered by the reduced outputs $y^k_N$ and (III) the difference between two reduced basis Kalman filters steered by the reduced outputs $y^k_N$ and the true outputs $y^k_h$ respectively.

Now, to quantify the quality of our RBKF estimation, we aim to find a statement about the estimation error of the output of interest $e^k_o := o(z^k_h - \hat{z}^k_N(y_h))$. As $e^k_o$ is a normally distributed random function in $Z_h$, we seek information about the expectation and the variance of $e^k_o$. The following Lemma shows, how the expectation and the variance of this estimation error can each be decomposed into four components.

**Lemma 4.1.** *Let $e^k_o = o(z^k_h - \hat{z}^k_N(y_h))$ be the RBKF estimation error of the output of interest. We define the four components*

$$e^k_{o,1} := o\left(z^k_h - z^k_N\right) - \sum_{i=1}^{k} \mathrm{Res}^i\left(z^{K-k+1}_{N,\mathrm{du},o}\right), \qquad e^k_{o,2} := o\left(z^k_N - \hat{z}^k_N(y_N)\right)$$

$$e^k_{o,3} := o\left(\hat{z}^k_N(y_N) - \hat{z}^k_N(y_h)\right), \qquad e^k_{o,4} := \sum_{i=1}^{k} \mathrm{Res}^i\left(z^{K-k+1}_{N,\mathrm{du},o}\right)$$

*for all $k \in \mathbb{K}$. Then the expectation of the estimation error can be decomposed into*

$$\mathrm{E}\left[e^k_o\right] = \mathrm{E}\left[e^k_{o,1}\right] + \mathrm{E}\left[e^k_{o,2}\right] + \mathrm{E}\left[e^k_{o,3}\right] + \mathrm{E}\left[e^k_{o,4}\right] \tag{4.15}$$

*and the variance of the estimation error can be bounded by*

$$\text{Var}\left(e_o^k\right) \leq \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \rho_3^k\right)\text{Var}\left(e_{o,1}^k\right) + \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \frac{1}{\rho_3^k}\right)\text{Var}\left(e_{o,2}^k\right)$$
$$+ \left(1 + \rho_1^k\right)\left(1 + \frac{1}{\rho_2^k}\right)\text{Var}\left(e_{o,3}^k\right) + \left(1 + \frac{1}{\rho_1^k}\right)\text{Var}\left(e_{o,4}^k\right) \tag{4.16}$$

*for any $\rho_1^k, \rho_2^k, \rho_3^k > 0$ and for all $k \in \mathbb{K}$.*

*Proof.* Due to the linearity of the output of interest, it is obvious that $e_o^k = e_{o,1}^k + e_{o,2}^k + e_{o,3}^k + e_{o,4}^k$. From the linearity of the expectation operator $\mathrm{E}\left[\cdot\right]$ we conclude (4.15). For deriving the variance bound, we recall that if $a, b \in \mathbb{R}$ then

$$(a + b)^2 \leq (1 + \rho)a^2 + \left(1 + \frac{1}{\rho}\right)b^2 \tag{4.17}$$

for any $\rho > 0$ (see [17] for a proof). Recalling $\mathring{z} := z - \mathrm{E}\left[z\right]$ the variance is reformulated as

$$\text{Var}\left(e_o^k\right) = \mathrm{E}\left[\left(e_o^k - \mathrm{E}\left[e_o^k\right]\right)^2\right]$$
$$= \mathrm{E}\left[\left(e_{o,1}^k - \mathrm{E}\left[e_{o,1}^k\right] + e_{o,2}^k - \mathrm{E}\left[e_{o,2}^k\right] + e_{o,3}^k - \mathrm{E}\left[e_{o,3}^k\right] + e_{o,4}^k - \mathrm{E}\left[e_{o,4}^k\right]\right)^2\right]$$
$$= \mathrm{E}\left[\left(\mathring{e}_{o,1}^k + \mathring{e}_{o,2}^k + \mathring{e}_{o,3}^k + \mathring{e}_{o,4}^k\right)^2\right]. \tag{4.18}$$

Then, applying the identity (4.17) successively three times with coefficients $\rho_1^k, \rho_2^k, \rho_3^k > 0$ we find

$$\left(\mathring{e}_{o,1}^k + \mathring{e}_{o,2}^k + \mathring{e}_{o,3}^k + \mathring{e}_{o,4}^k\right)^2 \leq \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \rho_3^k\right)\left(\mathring{e}_{o,1}^k\right)^2$$
$$+ \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \frac{1}{\rho_3^k}\right)\left(\mathring{e}_{o,2}^k\right)^2$$
$$+ \left(1 + \rho_1^k\right)\left(1 + \frac{1}{\rho_2^k}\right)\left(\mathring{e}_{o,3}^k\right)^2 + \left(1 + \frac{1}{\rho_1^k}\right)\left(\mathring{e}_{o,4}^k\right)^2. \tag{4.19}$$

Putting (4.19) in (4.18) and using the linearity of the expectation results in

$$\text{Var}\left(e_o^k\right) \leq \mathrm{E}\left[\left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \rho_3^k\right)\left(\mathring{e}_{o,1}^k\right)^2 + \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \frac{1}{\rho_3^k}\right)\left(\mathring{e}_{o,2}^k\right)^2\right.$$
$$+ \left(1 + \rho_1^k\right)\left(1 + \frac{1}{\rho_2^k}\right)\left(\mathring{e}_{o,3}^k\right)^2 + \left.\left(1 + \frac{1}{\rho_1^k}\right)\left(\mathring{e}_{o,4}^k\right)^2\right]$$
$$= \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \rho_3^k\right)\mathrm{E}\left[\left(\mathring{e}_{o,1}^k\right)^2\right] + \left(1 + \rho_1^k\right)\left(1 + \rho_2^k\right)\left(1 + \frac{1}{\rho_3^k}\right)\mathrm{E}\left[\left(\mathring{e}_{o,2}^k\right)^2\right]$$
$$+ \left(1 + \rho_1^k\right)\left(1 + \frac{1}{\rho_2^k}\right)\mathrm{E}\left[\left(\mathring{e}_{o,3}^k\right)^2\right] + \left(1 + \frac{1}{\rho_1^k}\right)\mathrm{E}\left[\left(\mathring{e}_{o,4}^k\right)^2\right].$$

With the definitions of $\mathring{e}_{o,1}^k, \mathring{e}_{o,2}^k, \mathring{e}_{o,3}^k$ and $\mathring{e}_{o,4}^k$ this leads directly to the statement (4.16). $\square$

Having this error decomposition, we will derive in the following sections error bounds for the expectations and variances of all four error components.

## 4.3. Error estimation for a noisy reduced basis model (Component $e_{o,1}^k$)

While the approximation error $e_z^k := z_h^k - z_N^k$ between the true solution of a stochastically forced PDE and its approximation *via* a reduced basis model is a crucial ingredient in the error estimation of the RB Kalman filter, it can be regarded independently from the Kalman filter problem setting and could be of individual interest.

We start with deriving the equations describing the evolution of the true expected solution $\bar{z}_h^k := \mathrm{E}\left[z_h^k\right]$ and the reduced expected solution $\bar{z}_N^k := \mathrm{E}\left[z_N^k\right]$. By taking the expectation on both sides of (3.1) and (4.1), due to the linearity of all terms and as $\mathrm{E}\left[\eta^k\right] = 0$ for all $k = 0, \ldots, K$ (white noise assumption) we obtain

$$a\left(\bar{z}_h^k, v; \boldsymbol{\mu}\right) = m\left(\bar{z}_h^{k-1}, v\ ; \boldsymbol{\mu}\right) + \Delta t\, b(v)u^k \quad \forall v \in Z_h \tag{4.20}$$

$$a\left(\bar{z}_N^k, v_N; \boldsymbol{\mu}\right) = m\left(\bar{z}_N^{k-1}, v_N\ ; \boldsymbol{\mu}\right) + \Delta t\, b(v_N)u^k \quad \forall v_N \in Z_N \tag{4.21}$$

for all $k \in \mathbb{K}$ with initial conditions $\bar{z}_h^0 = \mathrm{E}\left[z_h^0\right]$ and $\bar{z}_N^0 = \mathrm{E}\left[z_N^0\right]$. Note, that the solutions $\bar{z}_h^k$ and $\bar{z}_N^k$ are non-random functions.

Next, the evolution schemes for the unbiased true solution $\mathring{z}_h^k := z_h^k - \bar{z}_h^k$ and unbiased reduced solution $\mathring{z}_N^k := z_N^k - \bar{z}_N^k$ are obtained by subtracting (4.20) from (3.1) and (4.21) from (4.1) resulting in

$$a\left(\mathring{z}_h^k, v; \boldsymbol{\mu}\right) = m\left(\mathring{z}_h^{k-1}, v\ ; \boldsymbol{\mu}\right) + \left(\eta^k, v\right) \quad \forall v \in Z_h \tag{4.22}$$

$$a\left(\mathring{z}_N^k, v_N; \boldsymbol{\mu}\right) = m\left(\mathring{z}_N^{k-1}, v_N\ ; \boldsymbol{\mu}\right) + \left(\eta^k, v_N\right) \quad \forall v_N \in Z_N \tag{4.23}$$

for all $k \in \mathbb{K}$ with initial conditions $\mathring{z}_h^0 = z_h^0 - \bar{z}_h^0$ and $\mathring{z}_N^0 = z_N^0 - \bar{z}_N^0$. Furthermore, we define the mean residual and the unbiased residual respectively as

$$\overline{\mathrm{Res}}^k(v) := \mathrm{E}\left[\mathrm{Res}^k(v)\right] = -a\left(\bar{z}_N^k, v; \boldsymbol{\mu}\right) + m\left(\bar{z}_N^{k-1}, v\ ; \boldsymbol{\mu}\right) + \Delta t\, b(v)u^k \tag{4.24}$$

$$\mathring{\mathrm{Res}}^k(v) := \mathrm{Res}^k(v) - \mathrm{E}\left[\mathrm{Res}^k(v)\right] = -a\left(\mathring{z}_N^k, v; \boldsymbol{\mu}\right) + m\left(\mathring{z}_N^{k-1}, v\ ; \boldsymbol{\mu}\right) + \left(\eta^k, v\right) \tag{4.25}$$

and their Riesz representations $\left(\bar{r}^k, v\right) := \overline{\mathrm{Res}}^k(v)$, $\left(\mathring{r}^k, v\right) := \mathring{\mathrm{Res}}^k(v)$ for all $v \in Z_h$. Again, we omit the parameter dependence of the residuals for better readability.

### 4.3.1. Expectation of the approximation error

We will first derive error bounds for the expectation of the solution approximation error and subsequently derive a primal dual error bound for the expectation of the output approximation error.

**Lemma 4.2.** *Assuming the initial error to be $e_z^0 = 0$, the energy norm of the expected value of the approximation error $e_z^k$ is bounded by*

$$|||\mathrm{E}\left[e_z^k\right]||| \leq \Delta_{RB,E}^k := \left(\frac{1}{\Delta t\, \alpha_0} \sum_{i=1}^K \left\|\bar{r}^i\right\|^2\right)^{\frac{1}{2}} \tag{4.26}$$

*for all $k \in \mathbb{K}$.*

*Proof.* By subtracting (4.21) from (4.20) we find the error evolution

$$m\left(\bar{e}_z^k, v\ ; \boldsymbol{\mu}\right) + \Delta t\, a_0\left(\bar{e}_z^k, v\ ; \boldsymbol{\mu}\right) = m\left(\bar{e}_z^{k-1}, v\ ; \boldsymbol{\mu}\right) + \overline{\mathrm{Res}}^k(v)$$

for the mean approximation error $\bar{e}_z^k := \mathrm{E}\left[e_z^k\right]$. As this error evolution does not contain any random quantities, but merely deterministic quantities, the proof can be continued in analogy to ([25], Prop. 4.1) (while considering the slightly different definition for the residual). $\square$

**Remark 4.3.** We observe, that the expected value of the approximation error of a noisy linear system equals the RB approximation error of the respective linear system without noise. There exist various other approaches to bound this RB error, *e.g.* for general evolution equations [28] or in a space-time formulation [53], which could also be applied here.

In this Lemma we assumed the initial error to be zero. This requirement can be fulfilled by including the initial condition $\bar{z}_h^0$ in the reduced space $Z_N$ during the basis generation. In case that the initial error can not be set to zero, *e.g.* if the initial condition is parameter dependent, an upper bound for $\bar{e}^0$ has to be added to (4.26). This issue reappears when considering the error bound for the dual solutions. Here, the final condition is parameter dependent *via* the parameter dependent bilinear form $m(\cdot,\cdot\,;\boldsymbol{\mu})$ (*cf.* (4.8)). Hence, we cannot assume that the final dual approximation error $e_{\mathrm{du}}^{K+1} := z_{\mathrm{du}}^{K+1} - z_{N,\mathrm{du}}^{K+1}$ is zero. Therefore, the following lemma will provide an upper bound for the final dual error. But first, we introduce a new scalar product $(v,w)_m := m(v,w;\boldsymbol{\mu}^{**})$ and norm $\|v\|_m^2 = (v,v)_m$ based on the bilinear form $m$ and a second reference parameter $\boldsymbol{\mu}^{**}$. With respect to this norm the bilinear form $m$ is assumed to be coercive with coercivity constant $\bar{\sigma} \leq \sigma(\boldsymbol{\mu}) := \inf_{v \in Z} m(v,v\,;\boldsymbol{\mu}) / \|v\|_m^2$.

**Lemma 4.4.** *The final approximation errors* $e_{\mathrm{du}}^{K+1}(\boldsymbol{\mu}) := z_{\mathrm{du}}^{K+1}(\boldsymbol{\mu}) - z_{N,\mathrm{du}}^{K+1}(\boldsymbol{\mu})$ *and* $e_{\mathrm{du},o}^{K+1}(\boldsymbol{\mu}) := z_{\mathrm{du},o}^{K+1}(\boldsymbol{\mu}) - z_{N,\mathrm{du},o}^{K+1}(\boldsymbol{\mu})$ *are bounded by*

$$\left\|e_{\mathrm{du}}^{K+1}(\boldsymbol{\mu})\right\|_m \leq \Delta_{\mathrm{du,final}}(\boldsymbol{\mu}) := \frac{\|r_{\mathrm{final}}(\boldsymbol{\mu})\|_m}{\bar{\sigma}}$$

*and*

$$\left\|e_{\mathrm{du},o}^{K+1}(\boldsymbol{\mu})\right\|_m \leq \Delta_{\mathrm{du,final},o}(\boldsymbol{\mu}) := \frac{\|r_{\mathrm{final},o}(\boldsymbol{\mu})\|_m}{\bar{\sigma}}$$

*where the Riesz representations of the final residuals are defined as*

$$(r_{\mathrm{final}}(\boldsymbol{\mu}),v)_m := s(v) - m\left(z_{N,\mathrm{du}}^{K+1},v\,;\boldsymbol{\mu}\right) \quad \text{and} \quad (r_{\mathrm{final},o}(\boldsymbol{\mu}),w)_m := o(w) - m\left(z_{N,\mathrm{du},o}^{K+1},w\,;\boldsymbol{\mu}\right)$$

*for all* $v \in Z_{\mathrm{du}}$ *and* $w \in Z_{\mathrm{du},o}$.

*Proof.* With (4.8) we obtain for the final error

$$m\left(e_{\mathrm{du}}^{K+1}(\boldsymbol{\mu}),v\,;\boldsymbol{\mu}\right) = s(v) - m\left(z_{N,\mathrm{du}}^{K+1},v\,;\boldsymbol{\mu}\right) = (r_{\mathrm{final}}(\boldsymbol{\mu}),v)_m.$$

Setting $v = e_{\mathrm{du}}^{K+1}(\boldsymbol{\mu})$, invoking the coercivity $\bar{\sigma}$ of $m$ and applying Cauchy–Schwarz results in the expression for the bound. Similar arguments hold for the output of interest. $\qquad\square$

Note, in case of the bilinear form $m(\cdot,\cdot)$ being parameter independent, the final solution can be added to the reduced spaces, so that $z_{\mathrm{du}}^{K+1} \in Z_{N,\mathrm{du}}$ and $z_{\mathrm{du},o}^{K+1} \in Z_{N,\mathrm{du},o}$ resulting in zero final errors $e_{\mathrm{du}}^{K+1} = 0$ and $e_{\mathrm{du},o}^{K+1} = 0$.

**Lemma 4.5.** *The energy norm of the dual approximation errors* $e_{\mathrm{du}}^k(\boldsymbol{\mu}) := z_{\mathrm{du}}^k(\boldsymbol{\mu}) - z_{N,\mathrm{du}}^k(\boldsymbol{\mu})$ *and* $e_{\mathrm{du},o}^k(\boldsymbol{\mu}) := z_{\mathrm{du},o}^k(\boldsymbol{\mu}) - z_{N,\mathrm{du},o}^k(\boldsymbol{\mu})$ *is bounded by*

$$|||e_{\mathrm{du}}^k(\boldsymbol{\mu})|||_{\mathrm{du}} \leq \Delta_{\mathrm{du}}^k(\boldsymbol{\mu}) := \left(\frac{1}{\bar{\alpha}_0 \Delta t}\sum_{i=k}^K \left\|r_{\mathrm{du}}^i\right\|^2 + \bar{\sigma}\Delta_{\mathrm{du,final}}(\boldsymbol{\mu})^2\right)^{\frac{1}{2}},$$

$$|||e_{\mathrm{du},o}^k(\boldsymbol{\mu})|||_{\mathrm{du}} \leq \Delta_{\mathrm{du},o}^k(\boldsymbol{\mu}) := \left(\frac{1}{\bar{\alpha}_0 \Delta t}\sum_{i=k}^K \left\|r_{\mathrm{du},o}^i\right\|^2 + \bar{\sigma}\Delta_{\mathrm{du,final},o}(\boldsymbol{\mu})^2\right)^{\frac{1}{2}}$$

*for all* $k \in \mathbb{K}$ *where the residuals' Riesz representations are defined as* $(r_{\mathrm{du}}^i,v) := \mathrm{Res}_{\mathrm{du}}^k(v)$ *and* $\left(r_{\mathrm{du},o}^i,w\right) := \mathrm{Res}_{\mathrm{du},o}^k(w)$ *for all* $v \in Z_{\mathrm{du}}$ *and* $w \in Z_{\mathrm{du},o}$.

*Proof.* The proof can be found in [25].  □

Next, we define the modified reduced measurement output

$$
y_N^{*,k} := y_N^k + \sum_{i=1}^{k} \mathrm{Res}^i \left( z_{N,\mathrm{du}}^{K-k+i} \right) \tag{4.27}
$$

and derive an error bound for the expectation between this and the true measurement output.

**Proposition 4.6.** *The expected value of the error in the measurement output is bounded by*

$$
|\mathrm{E}\left[ y_h^k - y_N^{*,k} \right]| \le \Delta_{\mathrm{E}[y]}^k(\boldsymbol{\mu}) := \Delta_{RB,E}^k(\boldsymbol{\mu})\Delta_{\mathrm{du}}^k(\boldsymbol{\mu}) \tag{4.28}
$$

*for all $k \in \mathbb{K}$ and the expected value of the error in the modified output of interest, being the expectation of the first error component $e_{o,1}^k$, is bounded for all $k \in \mathbb{K}$ by*

$$
|\mathrm{E}\left[ e_{o,1}^k \right]| \le \Delta_{\mathrm{E}[e_{o1}]}^k(\boldsymbol{\mu}) := \Delta_{RB,E}^k(\boldsymbol{\mu})\Delta_{\mathrm{du},o}^k(\boldsymbol{\mu}) \tag{4.29}
$$

*where $\Delta_{RB,E}^k(\boldsymbol{\mu})$, $\Delta_{\mathrm{du}}^k(\boldsymbol{\mu})$ and $\Delta_{\mathrm{du},o}^k(\boldsymbol{\mu})$ are the error bounds from Lemma 4.2 and Lemma 4.5.*

*Proof.* For the expectation of the difference in the measurement outputs we find

$$
\mathrm{E}\left[ y_h^k - y_N^{*,k} \right] = \mathrm{E}\left[ y^k \right] - \mathrm{E}\left[ y_N^{*,k} \right] = s(\bar{z}_h^k) - s(\bar{z}_N^k) - \sum_{i=1}^{k} \bar{\mathrm{Res}}^i(z_{N,\mathrm{du}}^{K-k+i})
$$

where all terms are non random quantities. The same holds for

$$
\mathrm{E}\left[ o(z_h^k) - o(z_N^k) - \sum_{i=1}^{k} \mathrm{Res}^i(z_{N,\mathrm{du},o}^{K-k+i}) \right] = o(\bar{z}_h^k) - o(\bar{z}_N^k) - \sum_{i=1}^{k} \bar{\mathrm{Res}}^i(z_{N,\mathrm{du},o}^{K-k+i})
$$

From here, we refer to the proof from ([25], Prop. 4.4) for concluding the statement.  □

Although the whole error estimation procedure seems to be quite excessive, it will turn out in Section 4.4 that this error bound for the measurement output plays a crucial rule in the complete Kalman filter error estimation and that we have an interest in keeping this error bound tight.

One result of this subsection is, that in case of systems disturbed by white noise, the approximation error behaves in the mean exactly like the approximation error of a system without any noise.

*4.3.2. Variance of the approximation error*

In order to derive an error estimator for the variance of the approximation error, information about the covariance of the reduced solution is needed. Plugging the ansatz for the unbiased reduced solution $\mathring{z}_N^k := \sum_{i=1}^{N}(\mathring{z}_N^k)_i \varphi_i$ with coefficient vector $\mathring{\boldsymbol{z}}_N^k \in \mathbb{R}^N$ in (4.23) results in

$$
\boldsymbol{A}_N \mathring{\boldsymbol{z}}_N^k = \boldsymbol{M}_N \mathring{\boldsymbol{z}}_N^{k-1} + \boldsymbol{\eta}_N^k \tag{4.30}
$$

for all $k \in \mathbb{K}$ and with $\boldsymbol{\eta}_N^k$ defined element-wise through $(\boldsymbol{\eta}_N^k)_i := \left( \eta^k, \varphi_i \right)$ for all $i = 1, \ldots, N$. Now, with the definition of the auto-covariance matrix $\mathrm{cov}(\boldsymbol{z}_N^k) := \boldsymbol{Q}_N^k := \mathrm{E}\left[ \mathring{\boldsymbol{z}}_N^k (\mathring{\boldsymbol{z}}_N^k)^T \right]$, we can derive the evolution scheme for the reduced solution's autocovariance

$$
\boldsymbol{A}_N \boldsymbol{Q}_N^k \boldsymbol{A}_N^T = \boldsymbol{M}_N \boldsymbol{Q}_N^{k-1} \boldsymbol{M}_N^T + \boldsymbol{R}_{\eta\eta,N} \tag{4.31}
$$

with $\boldsymbol{R}_{\eta\eta,N} \in \mathbb{R}^{N \times N}$, $(\boldsymbol{R}_{\eta\eta,N})_{i,j} := (\varphi_i, R_{\eta\eta}\varphi_j)$ for all $i, j = 1, \ldots, N$. This evolution scheme is low dimensional, so that the covariance matrix of the reduced solution can be calculated online.

In analogy to the error estimator for the expectation, we will need a residual norm for bounding the variance, namely the expression $\mathrm{E}[\|\mathring{r}^k\|^2]$, where $\mathring{r}^k$ is the Riesz representation of the unbiased residual $\mathring{\mathrm{Res}}^k$. The following Lemma shows how to rapidly calculate this value online.

**Lemma 4.7.** *The expectation of the squared norm of the unbiased residual* $\mathrm{E}\left[\|\mathring{r}^k\|^2\right]$ *can be calculated* via

$$\mathrm{E}\left[\|\mathring{r}^k\|^2\right] = \mathrm{tr}\left(\boldsymbol{Q}_N^k \boldsymbol{K}_I\right) + \mathrm{tr}\left(\boldsymbol{Q}_N^{k-1} \boldsymbol{K}_{II}\right) + \mathrm{trace}_{Z_h}\left(R_{\eta\eta}\right)$$
$$-\mathrm{tr}\left(\boldsymbol{A}_N^{-1} \boldsymbol{M}_N \boldsymbol{Q}_N^{k-1} \boldsymbol{K}_{III}\right) - \mathrm{tr}\left(\boldsymbol{A}_N^{-1} \boldsymbol{K}_{\eta A}\right) \tag{4.32}$$

*for all* $k \in \mathbb{K}$ *with low dimensional matrices* $(\boldsymbol{K}_I)_{i,j} := (A\varphi_i, A\varphi_j)$, $(\boldsymbol{K}_{II})_{i,j} := (M\varphi_i, M\varphi_j)$, $(\boldsymbol{K}_{III})_{i,j} := (A\varphi_i, M\varphi_j)$, $(\boldsymbol{K}_{\eta A})_{i,j} = (\varphi_i, R_{\eta\eta}A\varphi_j)$ *for all* $i, j = 1, \ldots, N$.

*Proof.* The proof can be found in Appendix A.3 □

We can now derive an upper bound for the variance of the approximation errors in both outputs.

**Proposition 4.8.** *Assuming the initial error* $e^0 = z_h^0 - z_N^0$ *to be zero, the variance of the approximation error of the measurement output is bounded by*

$$\mathrm{Var}\left(y_h^k - y_N^{*,k}\right) \leq \Delta_{\mathrm{Var}(y)}^k := \frac{1}{\bar{\alpha}_0 \Delta t} \left(\sum_{i=1}^k \mathrm{E}\left[\|\mathring{r}^2\|^2\right]\right) \left(\Delta_{\mathrm{du}}^{K-k+1}\right)^2 \tag{4.33}$$

*and the approximation error of the first error component is bounded by*

$$\mathrm{Var}\left(e_{o,1}^k\right) \leq \Delta_{\mathrm{Var}(eo1)}^k := \frac{1}{\bar{\alpha}_0 \Delta t} \left(\sum_{i=1}^k \mathrm{E}\left[\|\mathring{r}^2\|^2\right]\right) \left(\Delta_{\mathrm{du},o}^{K-k+1}\right)^2. \tag{4.34}$$

*The sum* $\sum_{i=1}^k \mathrm{E}\left[\|\mathring{r}^2\|^2\right]$ *is rapidly computable as shown in Lemma* 4.7.

*Proof.* The proof can be found in Appendix A.4. □

In case, that the initial value is also a random function, an additional term including the initial covariance $\mathrm{cov}(\mathring{e}^0)$ has to be added.

In this section we derived error bounds to quantify the quality of reduced basis approximations of linear systems disturbed by white noise. The primal-dual technique provides us with sharp error bounds for linear outputs. In Section 4.4, those error bounds will play a crucial role in deriving an overall error estimator for the RBKF.

## 4.4. Error bounds for the deviation (components $e_{o,2}^k$, $e_{o,3}^k$, $e_{o,4}^k$)

In this subsection we will derive error bounds for the last three error components $e_{o,2}^k$, $e_{o,3}^k$ and $e_{o,4}^k$ of Lemma 4.1. As the solutions $z_N^k$, $\hat{z}_N^k(y_h)$ and $\hat{z}_N^k(y_N)$ – appearing in the two error components $e_{o,2}^k = o(z_N^k - \hat{z}_N^k(y_N))$ and $e_{o,3}^k = o(\hat{z}_N^k(y_h) - \hat{z}_N^k(y_N))$ – are all elements of the low dimensional space $Z_N$, we can expect them to be rapidly computable.

*4.4.1. The optimal reduced order filter $(e_{o,2}^k)$*

The second error component $e_{o,2}^k = o(z_N^k - \hat{z}_N^k(y_N))$ contains the difference between the reduced system's state $z_N^k$ and the reduced order Kalman filter estimation $\hat{z}_N^k(y_N)$ steered by the reduced system's outputs. As stated in Remark 3.1, $\hat{z}_N^k(y_N)$ is an optimal filter estimation for $z_N^k$. Consequently, statements about the estimation error are well known and are summarized in the following Lemma. In the following, we denote by $\boldsymbol{C}_o \in \mathbb{R}^N$ the vector evaluating the output of interest, defined as $(\boldsymbol{C}_o)_i := o(\varphi_i)$ for all $i = 1, \ldots, N$.

**Lemma 4.9.** *The expected value of the second error component equals zero*

$$\mathrm{E}\left[e_{o,2}^k\right] = \mathrm{E}\left[z_N^k - \hat{z}_N^k(y_N)\right] = 0 \tag{4.35}$$

*and the variance of the second error component is given by*

$$\mathrm{Var}\left(e_{o,2}^k\right) = \boldsymbol{C}_o^T \boldsymbol{P}_N^k \boldsymbol{C}_o. \tag{4.36}$$

*Proof.* The statement (4.35) results from standard Kalman filter analysis (*cf.* [31]). For (4.36) consider (2.1) and the fact that $\boldsymbol{P}_N^k = \mathrm{cov}(\boldsymbol{z}_N^k - \hat{\boldsymbol{z}}_N^k)$. □

Note, that all quantities in (4.35) and (4.36) are independent of the high dimensional approximation space $Z_h$ and can be calculated online.

*4.4.2. Expectation of the third error component $e_{o,3}^k$*

Now we want to address the third error component $e_{o,3}^k = o(\hat{z}_N^k(y_h) - \hat{z}_N^k(y_N))$, the deviation in the output of interest between the two differently steered Kalman filters. Therefor we first derive the error evolution scheme for the error $e_N^k := \hat{z}_N^k(y_h) - \hat{z}_N^k(y_N)$. Note, that $e_N^k \in Z_N$. By subtracting (4.4) from (3.4) and defining $\boldsymbol{G}_N := \boldsymbol{A}_N^{-1}(\boldsymbol{M}_N - \boldsymbol{L}_N^k \boldsymbol{C}_N^T \boldsymbol{A}_N^{-1} \boldsymbol{M}_N)$ we obtain

$$\boldsymbol{e}_N^k = \boldsymbol{G}_N \boldsymbol{e}_N^{k-1} + \boldsymbol{A}_N^{-1} \boldsymbol{L}_N^k \left(y_h^k - y_N^k\right) \tag{4.37}$$

where $\boldsymbol{e}_N^k \in \mathbb{R}^N$ is the coefficient vector, so that $e_N^k = \sum_{i=1}^N (\boldsymbol{e}_N^k)_i \varphi_i$. The initial error is $\boldsymbol{e}_N^0 = 0$ as $\hat{z}_N^0(y_h) = \hat{z}_N^0(y_N)$. To benefit from the primal dual error estimator, we rewrite (4.37) as

$$\boldsymbol{e}_N^k = \boldsymbol{G}_N \boldsymbol{e}_N^{k-1} + \boldsymbol{A}_N^{-1} \boldsymbol{L}_N^k \left(y_h^k - y_N^{*,k} + \sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right) \tag{4.38}$$

with $y_N^{*,k}$ the modified output from (4.27).

Obviously, this error system is stable (hence, the error is bounded), if the eigenvalues of $\boldsymbol{G}_N$ are inside the unit disc. We will learn in Section 4.6, that this is the case, if the reduced system is completely observable. To bound the error system (4.38), we derive an enveloping system. This enveloping system is found by a coordinate transformation and taking absolute values of all vector and matrix entries: for every matrix $\boldsymbol{G}_N$ an invertible matrix $\boldsymbol{T}$ can be found so that $\boldsymbol{D} := \boldsymbol{T}^{-1} \boldsymbol{G}_N \boldsymbol{T}$ is in Jordan normal form (*cf.* [30]). Then, we can rewrite (4.38) as:

$$\tilde{\boldsymbol{e}}_N^k = \boldsymbol{D} \tilde{\boldsymbol{e}}_N^{k-1} + \boldsymbol{T}^{-1} \boldsymbol{A}_N^{-1} \boldsymbol{L}_N^k \left(y_h^k - y_N^{*,k} + \sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right). \tag{4.39}$$

with $\tilde{\boldsymbol{e}}_N^k := \boldsymbol{T}^{-1} \boldsymbol{e}_N^k$. Now, we define the *enveloping system* of (4.38) as

$$\tilde{\boldsymbol{e}}_{\mathrm{N,env}}^k = |\boldsymbol{D}| \tilde{\boldsymbol{e}}_{\mathrm{N,env}}^{k-1} + |\tilde{\boldsymbol{B}}_N^k| \left| y_h^k - y_N^{*,k} + \sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right) \right| \tag{4.40}$$

with $\tilde{\boldsymbol{e}}_{\mathrm{N,env}}^0 = |\boldsymbol{T}^{-1}\boldsymbol{e}_N^0|$ and $\tilde{\boldsymbol{B}}_N^k := \boldsymbol{T}^{-1}\boldsymbol{A}_N^{-1}\boldsymbol{L}_N^k$. Note, that taking the absolute value of a matrix means, that we take the absolute values of all matrix entries, e.g. the absolute value of $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ is defined as $|\boldsymbol{D}| \in \mathbb{R}^{N\times N}$ with $(|\boldsymbol{D}|)_{i,j} = |(\boldsymbol{D})_{i,j}|$ for all $i, j = 1, \ldots, N$. We define the *enveloping solution* as $\boldsymbol{e}_{\mathrm{N,env}}^k := |\boldsymbol{T}|\tilde{\boldsymbol{e}}_{\mathrm{N,env}}^k$. Now, $\boldsymbol{e}_{\mathrm{N,env}}^k$ envelopes the error system in the sense that $\boldsymbol{e}_N^k \leq \boldsymbol{e}_{\mathrm{N,env}}^k$. Here (and in the following when used with vectors and matrices) "$\leq$" means an element-wise comparison, e.g. $(\boldsymbol{e}_N^k)_i \leq (\boldsymbol{e}_{\mathrm{N,env}}^k)_i$ for all $i = 1, \ldots, N$.

Using an enveloping system to bound the error $\boldsymbol{e}_N^k$ has one major advantage: if the eigenvalues of the original error system (4.38) are in the unit disc, then the eigenvalues of the enveloping system (4.40) will also be in the unit disc, and hence, the enveloping system will be stable.

Having found the enveloping system, we are now ready to derive an error bound for the expected value of the third component $e_{o,3}^k$ of the output estimation error (note, that writing in upper case after brackets or dashes symbolizes the raise to the power, e.g. $(\boldsymbol{D})^i$ or $|D|^i$, while simply writing in upper case signifies an indexing, e.g. $\boldsymbol{L}^i$).

**Proposition 4.10.** *The expected value of the error $e_{o,3}^k = o(\hat{z}_N^k(y_N) - \hat{z}_N^k(y_h))$ can be bounded by*

$$\mathrm{E}\left[e_{o,3}^k\right] \leq \Delta_{\mathrm{E}[e_{o3}]}^k \tag{4.41}$$

*for all $k \in \mathbb{K}$ with*

$$\Delta_{\mathrm{E}[e_{o3}]}^k := \left|\boldsymbol{C}_o^T\boldsymbol{T}\right|\left(|\boldsymbol{D}|^k\left|\boldsymbol{T}^{-1}\mathrm{E}\left[\boldsymbol{e}_N^0\right]\right| + \sum_{i=1}^k |\boldsymbol{D}|^{k-i}\left|\tilde{\boldsymbol{B}}_N^i\right|\left(\Delta_{\mathrm{E}[y]}^i + \left|\sum_{j=1}^i \bar{\mathrm{Res}}^j\left(z_{N,\mathrm{du}}^{K-i+j}\right)\right|\right)\right) \tag{4.42}$$

*with $\tilde{\boldsymbol{B}}_N^i = \boldsymbol{T}^{-1}\boldsymbol{A}_N^{-1}\boldsymbol{L}_N^i$ and with the measurement output error bound $\Delta_{\mathrm{E}[y]}^i$ from (4.28).*

*Proof.* In the following we define the expectation of the Kalman deviation error as $\bar{\boldsymbol{e}}_N^k := \mathrm{E}[\boldsymbol{e}_N^k]$ and the expectation of the outputs as $\bar{y_h}^k := \mathrm{E}[y_h^k]$ and $\bar{y}_N^{*,k} := \mathrm{E}[y_N^{*,k}]$. Now, we take the expectation on both sides of (4.38) and find

$$\bar{\boldsymbol{e}}_N^k = (\boldsymbol{G}_N)^k\bar{\boldsymbol{e}}_N^0 + \sum_{i=1}^k (\boldsymbol{G}_N)^{k-i}\boldsymbol{A}_N^{-1}\boldsymbol{L}_N^i\left(\bar{y}_h^i - \bar{y}_N^{*,i} + \sum_{j=1}^i \bar{\mathrm{Res}}^j\left(z_{N,\mathrm{du}}^{K-i+j}\right)\right)$$

$$= \boldsymbol{T}(\boldsymbol{D})^k\boldsymbol{T}^{-1}\bar{\boldsymbol{e}}_N^0 + \boldsymbol{T}\sum_{i=1}^k (\boldsymbol{D})^{k-i}\boldsymbol{T}^{-1}\boldsymbol{A}_N^{-1}\boldsymbol{L}_N^i\left(\bar{y}_h^i - \bar{y}_N^{*,i} + \sum_{j=1}^i \bar{\mathrm{Res}}^j\left(z_{N,\mathrm{du}}^{K-i+j}\right)\right). \tag{4.43}$$

Knowing that $e_{o,3}^k = \boldsymbol{C}_o^T\boldsymbol{e}_N^k$, we multiply (4.43) on both sides with $\boldsymbol{C}_o^T$ and derive the upper bound

$$\boldsymbol{C}_o^T\bar{\boldsymbol{e}}_N^k = \boldsymbol{C}_o^T\boldsymbol{T}\left((\boldsymbol{D})^k\boldsymbol{T}^{-1}\bar{\boldsymbol{e}}_N^0 + \sum_{i=1}^k (\boldsymbol{D})^{k-i}\boldsymbol{T}^{-1}\boldsymbol{A}_N^{-1}\boldsymbol{L}_N^i\left(\bar{y}^i - \bar{y}_N^{*,i} + \sum_{j=1}^i \bar{\mathrm{Res}}^j\left(z_{N,\mathrm{du}}^{K-i+j}\right)\right)\right)$$

$$\leq |\boldsymbol{C}_o^T\boldsymbol{T}|\left((|\boldsymbol{D}|)^k|\boldsymbol{T}^{-1}\bar{\boldsymbol{e}}_N^0| + \sum_{i=1}^k |\boldsymbol{D}|^{k-i}\left|\tilde{\boldsymbol{B}}_N^i\right|\left(\Delta_{\mathrm{E}[y]}^i + \left|\sum_{j=1}^i \bar{\mathrm{Res}}^j\left(z_{N,\mathrm{du}}^{K-i+j}\right)\right|\right)\right)$$

leading to the statement in (4.42). $\square$

Note, that although a sum of the residual sum appears in the error estimation, the error bound will turn out to be very sharp, cf. the experiments in Section 5.

*4.4.3. Variance of the third error component $e_{o,3}^k$*

Before deriving the variance of the third error component, we will have a look at some quantities helping us to describe and calculate this desired error variance. Firstly, we will extend the formulation of the autocovariance of the reduced solution $\boldsymbol{Q}_N^i$ to a general covariance $\boldsymbol{Q}_N^{i,j} := \mathrm{cov}(\boldsymbol{z}_N^i, \boldsymbol{z}_N^j)$. The following Lemma will help us to calculate these covariances.

**Lemma 4.11.** *Be $\boldsymbol{Q}_N^j$ the autocovariance given by* (4.31), *then the covariance $\boldsymbol{Q}_N^{i,j} = \mathrm{cov}(\boldsymbol{z}_N^i, \boldsymbol{z}_N^j)$ for $i > j$ is given by*

$$\boldsymbol{Q}_N^{i,j} = \left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-j}\boldsymbol{Q}_N^j. \tag{4.44}$$

*Proof.* For time indices $i > j$ the time evolution of the unbiased reduced solution in (4.30) can be reformulated to

$$\mathring{\boldsymbol{z}}_N^i = \left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-j}\mathring{\boldsymbol{z}}_N^j + \sum_{l=j+1}^{i}\left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-l}\boldsymbol{A}_N^{-1}\boldsymbol{\eta}_N^l. \tag{4.45}$$

Knowing that $\mathrm{cov}(\boldsymbol{\eta}^l, \boldsymbol{z}_N^j) = 0$ for $l > j$ (noise independent of former states) we find

$$\boldsymbol{Q}_N^{i,j} = \mathrm{E}\left[\mathring{\boldsymbol{z}}_N^i(\mathring{\boldsymbol{z}}_N^j)^T\right] = \mathrm{E}\left[\left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-j}\mathring{\boldsymbol{z}}_N^j\left(\mathring{\boldsymbol{z}}_N^j\right)^T\right].$$

Due to the linearity of the operators, this results in the equation stated in the lemma. $\square$

Note, that if $j > i$ then the solution is simply found by transposing $\boldsymbol{Q}_N^{j,i}$.

In the error evolution for $\boldsymbol{e}_N^k$ in (4.38) appears the residual sum $\sum_{i=1}^k \mathrm{Res}^i(z_{N_{\mathrm{du}}}^{K-k+i})$. It will be important to find a way to rapidly calculate this sum for the variance error estimator. Defining the coefficient vector $\boldsymbol{z}_{N,\mathrm{du}}^k \in \mathbb{R}^N$ via $(z_{N,\mathrm{du}}^k)_i := \left(z_{N,\mathrm{du}}^k, \varphi_{du,i}\right)$, we find

**Lemma 4.12.** *The variance of the residual sum $\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i(z_{N,\mathrm{du}}^{K-k+i})\right)$ is given by*

$$\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right) = \sum_{i=1}^k\sum_{j=1}^k\left(\boldsymbol{z}_{N,\mathrm{du}}^{K-k+i}\right)^T\boldsymbol{R}_{rr}^{i,j}\boldsymbol{z}_{N,\mathrm{du}}^{K-k+j} \tag{4.46}$$

*where the matrix $\boldsymbol{R}_{rr}^{i,j} \in \mathbb{R}^{N_{\mathrm{du}}\times N_{\mathrm{du}}}$ is given by*

$$\boldsymbol{R}_{rr}^{i,j} = \boldsymbol{K}_A\boldsymbol{Q}_N^{i,j}\boldsymbol{K}_A^T + \boldsymbol{K}_M\boldsymbol{Q}_N^{i-1,j-1}\boldsymbol{K}_M^T + \delta_{i,j}\boldsymbol{R}_{\eta\eta,\mathrm{du}} - \boldsymbol{K}_A\boldsymbol{Q}_N^{i,j-1}\boldsymbol{K}_M^T$$
$$-\boldsymbol{K}_A\boldsymbol{Q}_N^{j,i-1}\boldsymbol{K}_M^T - \boldsymbol{K}_A\boldsymbol{R}_{z\eta}^{i,j} - \left(\boldsymbol{R}_{z\eta}^{j,i}\right)^T\boldsymbol{K}_A^T + \boldsymbol{K}_M\boldsymbol{R}_{z\eta}^{i-1,j} + \left(\boldsymbol{R}_{z\eta}^{j-1,i}\right)^T\boldsymbol{K}_M^T \tag{4.47}$$

*with $\boldsymbol{Q}_N^{i,j}$ the covariance of the reduced solution from Lemma* 4.11, *the covariance matrix*

$$\boldsymbol{R}_{z\eta}^{i,j} = \begin{cases} \left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-j}\boldsymbol{A}_N^{-1}\boldsymbol{R}_{\eta\eta,\mathrm{du}} & i \geq j \\ \boldsymbol{0} & i < j \end{cases}$$

*and the matrices $\boldsymbol{K}_A, \boldsymbol{K}_M \in \mathbb{R}^{N_{\mathrm{du}}\times N}$ and $\boldsymbol{R}_{\eta\eta,\mathrm{du}} \in \mathbb{R}^{N\times N_{\mathrm{du}}}$ are defined by*

$$(\boldsymbol{K}_A)_{n,m} := (A\varphi_{\mathrm{du},n}, \varphi_m), (\boldsymbol{K}_M)_{n,m} := (M\varphi_{\mathrm{du},n}, \varphi_m), (\boldsymbol{R}_{\eta\eta,\mathrm{du}})_{m,n} := (\varphi_m, R_{\eta\eta}\varphi_{du,n})$$

*for all $n = 1, \ldots, N_{\mathrm{du}}$ and $m = 1, \ldots, N$.*

*Proof.* The proof can be found in Appendix A.5. $\square$

**Remark 4.13.** When treating problems with many time steps, calculating the expression in (4.46) can be computationally very expensive, as the computational cost rises quadratically with the number of time steps. In Section 4.5.1 we will propose a way to approximate the variance of the residual sum, which can accelerate the error estimation procedure drastically.

We can now derive a bound for the variance of the third error component. In the following, taking the square root of a vector means taking the square root of every entry of the vector.

**Proposition 4.14.** *The variance of the third error component* $\mathrm{Var}\left(e_{o,3}^k\right)$ *is for all* $k \in \mathbb{K}$ *bounded by*

$$\mathrm{Var}\left(e_{o,3}^k\right) \leq \Delta_{\mathrm{Var}(e_{o3})}^k(\boldsymbol{\mu}) := \left|\boldsymbol{C}_o^T\boldsymbol{T}\right|\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k\left|\boldsymbol{T}^T\boldsymbol{C}_o\right| \tag{4.48}$$

*where the matrix* $\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k \in \mathbb{R}^{N\times N}$ *is calculated iteratively* via

$$
\begin{aligned}
\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k =& |\boldsymbol{D}|\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\left|\boldsymbol{D}^T\right| + \left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu}) \\
&+ \left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right) + \sqrt{\Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu})}\left\{|\boldsymbol{D}|\sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)}\left|\tilde{\boldsymbol{B}}^T\right|\right. \\
&+ \left|\tilde{\boldsymbol{B}}\right|\sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)^T}\left|\boldsymbol{D}^T\right| + 2\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\sqrt{\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)}\right\} \\
&+ \sqrt{\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)}\left\{|\boldsymbol{D}|\sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)}\left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| + \left|\tilde{\boldsymbol{B}}^k\right|\sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)^T}\left|\boldsymbol{D}^T\right|\right\}
\end{aligned}
\tag{4.49}
$$

*with initial matrix* $\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^0 = |\boldsymbol{T}^{-1}\mathrm{E}\left[\mathring{\boldsymbol{e}}_N^0(\mathring{\boldsymbol{e}}_N^0)^T\right]\boldsymbol{T}^{-T}|$ *and the vector* $\tilde{\boldsymbol{B}}^k := \boldsymbol{T}^{-1}\boldsymbol{A}^{-1}\boldsymbol{L}_N^k$. *The variance of the residual sum* $\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i(z_{N,\mathrm{du}}^{K-k+i})\right)$ *can be calculated with Lemma* 4.12.

*Proof.* The proof can be found in Appendix A.6. □

### 4.4.4. Expectation and variance of the fourth error component $e_{o,4}^k$

To use the primal dual error estimator, we had to add the residual sum $\sum_{i=1}^k \mathrm{Res}(z_{N,\mathrm{du},o}^{K-k+i})$ to the reduced output of interest $o(z_N^k)$. Therefore, we must now in a final step elaborate an expression for the expectation and the variance of this residual sum.

**Lemma 4.15.** *The expectation of the fourth error component* $e_{o,4}^k = \sum_{i=1}^k \mathrm{Res}^i(z_{N,\mathrm{du},o}^{K-k+1})$ *is given by*

$$\mathrm{E}\left[e_{0,4}^k\right] = \sum_{i=1}^k \bar{\mathrm{Res}}^i\left(z_{N,\mathrm{du},o}^{K-k+i}\right) \tag{4.50}$$

*with the mean residual* $\bar{\mathrm{Res}}^i$ *from* (4.24).

*Proof.* As the dual solution $z_{N,\mathrm{du},o}^k$ is a non-random function and taking the expectation is a linear operation, finding (4.50) is straightforward. □

Next, we derive the variance of the fourth error component $e_{o,4}$.

**Lemma 4.16.** *The variance of the residual sum* $\mathrm{Var}\left(\sum_{i=1}^{k}\mathrm{Res}^i(z_{N,\mathrm{du},o}^{K-k+i})\right)$ *is given by*

$$\Delta_{Var(e_{o4})}^{k}(\boldsymbol{\mu}) := \mathrm{Var}\left(\sum_{i=1}^{k}\mathrm{Res}^i\left(z_{N,\mathrm{du},o}^{K-k+i}\right)\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(\boldsymbol{z}_{N,\mathrm{du},o}^{K-k+i}\right)^T \boldsymbol{R}_{rr,o}^{i,j}\boldsymbol{z}_{N,\mathrm{du},o}^{K-k+i} \tag{4.51}$$

*where the matrix* $\boldsymbol{R}_{rr,o}^{i,j} \in \mathbb{R}^{N_{\mathrm{du},o}\times N_{\mathrm{du},o}}$ *is given by*

$$\boldsymbol{R}_{rr,o}^{i,j} = \boldsymbol{K}_{A,o}\boldsymbol{Q}_N^{i,j}\boldsymbol{K}_{A,o}^T + \boldsymbol{K}_{M,o}\boldsymbol{Q}_N^{i-1,j-1}\boldsymbol{K}_{M,o}^T + \delta_{i,j}\boldsymbol{R}_{\eta\eta,\mathrm{du},o} - \boldsymbol{K}_{A,o}\boldsymbol{Q}_N^{i,j-1}\boldsymbol{K}_{M,o}^T$$
$$+\boldsymbol{K}_{A,o}\boldsymbol{Q}_N^{j,i-1}\boldsymbol{K}_{M,o}^T - \boldsymbol{K}_{A,o}\boldsymbol{R}_{z\eta,o}^{i,j} - \left(\boldsymbol{R}_{z\eta,o}^{j,i}\right)^T\boldsymbol{K}_{A,o}^T + \boldsymbol{K}_{M,o}\boldsymbol{R}_{z\eta,o}^{i-1,j} + \left(\boldsymbol{R}_{z\eta,o}^{j-1,i}\right)^T\boldsymbol{K}_{M,o}^T$$

*with* $\boldsymbol{Q}_N^{i,j}$ *the covariance of the reduced solution from Lemma 4.11, the covariance matrix*

$$\boldsymbol{R}_{z\eta,o}^{i,j} = \begin{cases} \left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-j}\boldsymbol{A}_N^{-1}\boldsymbol{R}_{\eta\eta,\mathrm{du},o} & i \geq j \\ \boldsymbol{0} & i < j \end{cases}$$

*and the matrices* $\boldsymbol{K}_{A,o}, \boldsymbol{K}_{M,o} \in \mathbb{R}^{N_{\mathrm{du},o}\times N}$ *and* $\boldsymbol{R}_{\eta\eta,\mathrm{du},o} \in \mathbb{R}^{N\times N_{\mathrm{du},o}}$ *defined by*

$$(\boldsymbol{K}_{A,o})_{n,m} := (A\varphi_{\mathrm{du},o,n}, \varphi_m), (\boldsymbol{K}_M)_{n,m} := (M\varphi_{\mathrm{du},o,n}, \varphi_m), (\boldsymbol{R}_{\eta\eta,\mathrm{du},o})_{m,n} := (\varphi_m, R_{\eta\eta}\varphi_{\mathrm{du},o,n})$$

*for all* $n = 1,\ldots,N_{\mathrm{du},o}$ *and* $m = 1,\ldots,N$.

*Proof.* The proof is similar to the proof of Lemma 4.12, only that the dual solution $z_{N,\mathrm{du},o}^k$ and the dual basis functions $\varphi_{\mathrm{du},o,i}$ of the output of interest are used instead of $z_{N,\mathrm{du}}^k$ and $\varphi_{\mathrm{du},o}^i$. $\qquad\square$

### 4.5. Complete RBKF output error estimation

Finally, we assemble all the upper bounds for the error components to derive an overall error bound for the RBKF estimation error.

**Proposition 4.17.** *The expectation and the variance of the RBKF output estimation error* $e_o^k = o(z_h^k - \hat{z}_N^k(y_h))$ *are bounded by*

$$\left|\mathrm{E}\left[e_o^k\right]\right| \leq \Delta_{\mathrm{E}[e_{o1}]}^k(\boldsymbol{\mu}) + \Delta_{\mathrm{E}[e_{o3}]}^k(\boldsymbol{\mu}) + \left|\sum_{i=1}^{k}\bar{\mathrm{Res}}^i\left(z_{N,\mathrm{du},o}^{K-k+i},\boldsymbol{\mu}\right)\right|$$

*and*

$$\mathrm{Var}\left(e_o^k\right) \leq \left(1+\rho_1^k\right)\left(1+\rho_2^k\right)\left(1+\rho_3^k\right)\Delta_{\mathrm{Var}(e_{o1})}^k(\boldsymbol{\mu}) + \left(1+\rho_1^k\right)\left(1+\rho_2^k\right)\left(1+\frac{1}{\rho_3^k}\right)\boldsymbol{C}_o\boldsymbol{P}_N^k(\boldsymbol{\mu})\boldsymbol{C}_o^T$$
$$+\left(1+\rho_1^k\right)\left(1+\frac{1}{\rho_2^k}\right)\Delta_{\mathrm{Var}(e_{o3})}^k(\boldsymbol{\mu}) + \left(1+\frac{1}{\rho_1^k}\right)\Delta_{\mathrm{Var}(e_{o4})}^k(\boldsymbol{\mu}) \tag{4.52}$$

*with optimal coefficients*

$$\rho_1^k := \sqrt{\frac{\Delta_{\mathrm{Var}(e_{o4})}^k}{\Delta_{\mathrm{Var}(e_{o1})}^k + \boldsymbol{C}_o^T\boldsymbol{P}_N^k\boldsymbol{C}_o + \Delta_{\mathrm{Var}(e_{o3})}^k}},$$

$$\rho_2^k := \sqrt{\frac{\Delta_{\mathrm{Var}(e_{o3})}^k}{\Delta_{\mathrm{Var}(e_{o1})}^k + \boldsymbol{C}_o^T\boldsymbol{P}_N^k\boldsymbol{C}_o}}, \quad \rho_3^k := \sqrt{\frac{\boldsymbol{C}_o^T\boldsymbol{P}_N^k\boldsymbol{C}_o}{\Delta_{\mathrm{Var}(e_{o1})}^k}}$$

*minimizing the sum* (4.52).

*Proof.* The stated bound for the expectation $|\mathrm{E}\left[e_o^k\right]|$ is a direct consequence of Lemma 4.1, Proposition 4.6, Lemma 4.9, Proposition 4.10 and Lemma 4.15.

For the variance bound we take the decomposition of $\mathrm{Var}\left(e_o^k\right)$ from Lemma 4.1 and bound (4.16) using $\mathrm{Var}\left(e_{o,1}(\boldsymbol{\mu})^k\right) \leq \Delta_{\mathrm{Var}(e_{o1})}^k(\boldsymbol{\mu})$ from Proposition 4.8, the exact value for $\mathrm{Var}\left(e_{o,2}^k(\boldsymbol{\mu})\right) = \boldsymbol{C}_o \boldsymbol{P}_N^k \boldsymbol{C}_o^T$ from Lemma 4.9, the bound for $\mathrm{Var}\left(e_{o,3}^k\right) \leq \Delta_{\mathrm{Var}(e_{o3})}^k$ from Proposition 4.14 and the bound $\mathrm{Var}\left(e_{0,4}^k\right) \leq \Delta_{\mathrm{Var}(e_{o4})}^k$ from Lemma 4.16. The optimal values for the coefficients $\rho_1, \rho_2$ and $\rho_3$ in Lemma 4.1 are found by a standard minimization. The function $f : \mathbb{R}^3 \to \mathbb{R}$

$$f(\rho_1^k, \rho_2^k, \rho_3^k) = \mathrm{R.H.S.}(4.52)$$

is a convex function for fixed positive values $\Delta_{\mathrm{Var}(e_{o1})}^k, \boldsymbol{C}_o \boldsymbol{P}_N^k \boldsymbol{C}_o^T, \Delta_{\mathrm{Var}(e_{o3})}^k, \Delta_{\mathrm{Var}(e_{o4})}^k$. The global minimum of $f$ can easily be found by a standard minimization. $\qquad\square$

### 4.5.1. Simplifications of the error estimation procedure

All online calculations in the error estimation procedures are independent of the dimension of the discretization space $Z_h$ (for the offline/online decomposition of the decomposition of the terms needed in the error estimation, *e.g.* $\overline{\mathrm{Res}}^i(z_{N,\mathrm{du}}^{K-k+i}, \boldsymbol{\mu})$, $\boldsymbol{K}_A(\boldsymbol{\mu})$, *etc.*, we refer to [25, 28]). However, in case of simulations involving many time steps $t^k$, the calculations of the reduced solution covariances $\boldsymbol{Q}_N^{i,j}$ (see Lem. 4.11) and the calculation of the covariances $\boldsymbol{R}_{rr}^{i,j}$ (see Lem. 4.12) for all combination of time step indices $i, j = 1, \ldots, K$ may be computationally expensive.

Although the calculation and incorporation of all those quantities is inevitable to find a rigorous error bound, one could content with a mathematically non-rigorous error estimator by neglecting the calculation of covariances $\boldsymbol{Q}_N^{i,j}$, $\boldsymbol{R}_{rr}^{i,j}$ for some specific time steps $|i - j| > \varepsilon_k$ and set them to zero, *cf.* [11] where a similar reduction technique has been applied earlier. This simplifies the calculations in (4.44) and (4.47) making the calculation effort of those quantities independent of the overall number of time steps. In the numerical experiments in Section 5 we will demonstrate, that calculating only a subset of covariances does not influence the value of the error estimator greatly in the chosen example.

### 4.6. Link between observability and stability of the RBKF

It is well known, that when performing Kalman filter estimations using an incorrect model, the mean square estimation error may become unbounded with time leading to an asymptotically unstable filter. This effect is also known as the divergence problem of the Kalman filter (*e.g.* [23, 46, 52]). As the RBKF uses only an approximative model, it could potentially turn out to be an unstable filter. It has been shown in previous works (*e.g.* [12, 23, 46]) under which conditions the divergence of the Kalman filter is bounded and that such a bound can be calculated. The error estimation procedure in Section 4 gives such an upper bound for the estimation error – *a posteriori* and for finite time. Now, we aim at finding an *a priori* statement about the filter stability.

In the following proposition we will show, that the stability of the filter depends on the complete observability of the reduced system. Note that the complete observability is not given *a priori* and, hence, has to be verified for every new reduced basis. The reduced system (4.1) with outputs $y_N^k$ from (4.2) is said to be completely observable, if for each time step $t^k$, with $k \in \mathbb{N}$ sufficiently large, there exists a positive integer $n$ such that the state $z_N^k$ can be uniquely determined by the sequence of observations $y_N^{k-n}, y_N^{k-n+1}, \ldots, y_N^k$. There exist various ways to verify the complete observability of discretized partial differential equations (see [12] for example).

**Proposition 4.18.** *Assume that the reduced basis approximation error is bounded by* $\left\| z_h^k - z_N^k \right\| < C_z$ *for all* $k \in \mathbb{N}$, *then the RBKF is stable (*i.e. $\left\| z_h^k - \hat{z}_N^k(y_h) \right\|$ *is bounded for all* $k \in \mathbb{N}$) *if and only if the reduced system* (4.1), (4.2)$_1$ *is uniformly completely observable.*

*Proof.* We can decompose the estimation error $z_h^k - \hat{z}_N^k(y_h)$ into three components, *cf.* (4.14). The norm of the first component $\left\| z_h^k - z_N^k \right\|$ is bounded by assumption. Hence, also the approximation error of the measurement output is bounded by $|y_h^k - y_N^k| \leq C_y$.
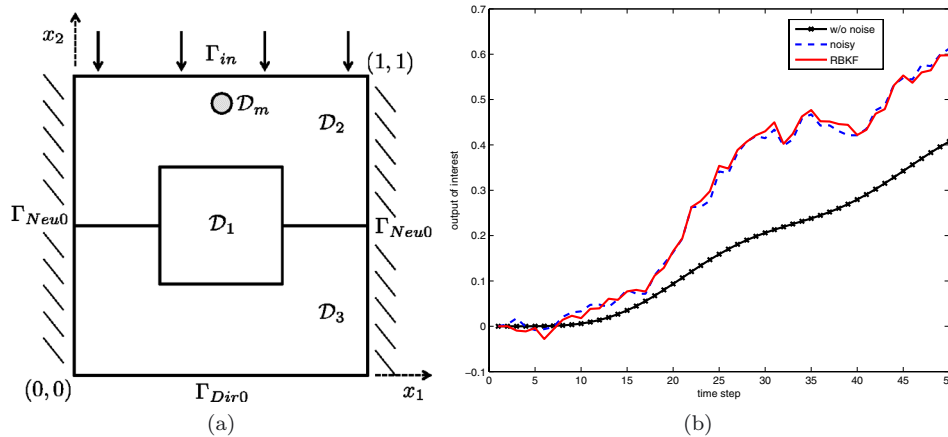
FIGURE 1. The problem setting is illustrated in (a). In (b) the output of interest of a system with parameters $(0.1, 10)$ without noise influence (*black*) and with an exemplary noise realization (*blue dashed line*) are plotted over time. Also the RBKF output estimate for that output (*red*) is plotted over time. (Color online).

The second component $z_N^k - \hat{z}_N^k(y_N)$ is the estimation error of a standard Kalman filter for a reduced system. We know (*e.g.* from [31], Thm. 7.4) that the Kalman filter is stable, if the reduced system is uniformly completely observable, which is given by assumption.

We recall, that the evolution of the third component $\hat{z}_N^k(y_N) - \hat{z}_N^k(y_h)$ is described by the system $e_N^k = G_N e_N^{k-1} + A_N^{-1} L_N^k (y_h^k - y_N^k)$ with $e_N^k = \hat{z}_N^k(y_N) - \hat{z}_N^k(y_h)$ (see (4.37)). We learn from [12] or ([31], Thm. 7.4), that if and only if the system (4.1) with outputs (4.2) is completely observable, then the eigenvalues of the matrix $G_N$ are inside the unit disc. Hence, the error system (4.37) is stable. Therefore, the norm of the third error component is bounded and its value depends on $C_y$. $\qquad\square$

Finally we give a short comment on the error estimation. As mentioned earlier, we are examining and bounding the error between the "true" noisy FEM solution and the RBKF estimation. When using the RBKF in practice, one would feed the measurements from a real system into the RBKF and not the FEM model output $y_h^k$. The estimation error would then, of course, be different because the FEM model does probably not capture the system's behavior in every detail and the real noise disturbing the system would differ from the assumed noise. However, the upper bound presented here still provides a reasonable error estimator and the stability of the RBKF is still given as it depends solely on the complete observability of the reduced system.

## 5. NUMERICAL EXPERIMENTS

We illustrate the performance of the RBKF with a numerical example. Therefor we consider the heat transfer in a plate consisting of different materials with different heat conductivities. The plate is assumed to be subject to a locally correlated noisy heat inflow on the whole domain (*e.g.* sun rays heating the plate) disturbing randomly the temperature distribution. We take temperature measurements in one spot of the plate, which are also corrupted by noise (*e.g.* because we bought cheap measurement instruments). The objective is to estimate the actual average temperature on a subdomain of the plate *via* this measurement data and to gain information on how reliable these estimates are.

Let $\boldsymbol{x} = (x_1, x_2)^T \in \mathbb{R}^2$ be a point on the plate being described by the domain $\mathcal{D} = [0, 1]^2$. As depicted in the left of Figure 1 the domain $\mathcal{D}$ is divided into three subdomains $\mathcal{D}_1 := [0.25, 0.75]^2$, $\mathcal{D}_2 := ([0, 1] \times [0.5, 1]) \setminus \mathcal{D}_1$ and $\mathcal{D}_3 := \mathcal{D} \setminus (\mathcal{D}_1 \cup \mathcal{D}_2)$. The three domains have different heat conductivities $\kappa(\boldsymbol{x}, \boldsymbol{\mu}) = 0.1$ for $\boldsymbol{x} \in \mathcal{D}_3$, $\kappa(\boldsymbol{x}, \boldsymbol{\mu}) = \mu_1$ for $\boldsymbol{x} \in \mathcal{D}_1$ and $\kappa(\boldsymbol{x}, \boldsymbol{\mu}) = \mu_2$ for $\boldsymbol{x} \in \mathcal{D}_2$ depending on the parameters $(\mu_1, \mu_2) = \boldsymbol{\mu}^T \in \mathcal{P} := [0.1, 10]^2$.

The time interval $[0, T]$ with $T = 1$ is divided into $K = 50$ equally spaced time steps and we are interested in the temperature distribution $z^k(\boldsymbol{\mu}) = z(t^k, \boldsymbol{\mu}) \in Z$ for all time steps $t^k$ in this interval. Here, $Z = Z(\mathcal{D}) := \{v | \int_{\mathcal{D}} v^2 + \int_{\mathcal{D}} \nabla v \cdot \nabla v < \infty\}$ and the scalar product is defined as $(v, w) := a_0(v, w; (1, 1))$. The plate has insulating boundaries with $\partial_{x_1} z^k(\boldsymbol{\mu}) = 0$ on $\Gamma_{Neu0} = \{x_1 = 0 \wedge 0 \leq x_2 \leq 1\} \cup \{x_1 = 1 \wedge 0 \leq x_2 \leq 1\}$, a cooling boundary with a constant temperature $z^k(\boldsymbol{\mu}) = 0$ on $\Gamma_{Dir0} := \{x_2 = 0 \wedge 0 \leq x_1 \leq 1\}$ and an oscillating heat inflow $\partial_{x_2} z^k(\boldsymbol{\mu}) = \sin^2(2\pi f t^k)$ on the boundary $\Gamma_{in} := \{x_2 = 1, 0 \leq x_1 \leq 1\}$ for all time indices $k = 0, \ldots, K$. We assume the non-random initial condition $z^0(\boldsymbol{\mu}) = 0$.

Modeling the heat transfer ($cf.$ [6]) in a time-discrete weak formulation, $z$ is solution to the stochastically forced PPDE (2.4) with bilinear forms $m(v, w; \boldsymbol{\mu}) = \int_{\mathcal{D}} vw$ and $a_0(v, w; \boldsymbol{\mu}) = \int_{\mathcal{D}} \kappa(\boldsymbol{\mu}, \boldsymbol{x}) \nabla v \cdot \nabla w$. The linear form is given by $b(v, \boldsymbol{\mu}) = \int_{\Gamma_{in}} v$ and the input is $u^k := \sin^2(2\pi f t^k)$ with frequency $f = 1$. Hence, the lower bound for the coercivity constant is $\bar{\alpha}_0 = 0.1$. Obviously, $m$ and $b$ are parameter independent. The bilinear form $a_0$ is parameter affine as required in (2.5) with coefficients $\Theta_a^1 = \mu_1$, $\Theta_a^2 = \mu_2$, $\Theta_a^3 = 0.1$ and components $a_0^q(v, w) = \int_{\mathcal{D}_q} \nabla v \cdot \nabla w$ with $q = 1, \ldots, 3$.

In this experiment we will generate synthetic noise influencing the temperature distribution on the plate. To that purpose the additive white noise stochastic process $\eta$ in (2.4) is defined $via$ the covariance operator $(R_{\eta\eta} w)(x_1) := \int_{\mathcal{D}} \mathcal{K}(x_1, x_2) w(x_2) dx_2$ with a kernel $\mathcal{K}(x_1, x_2) := \sigma_\eta^2 e^{-(x_1-x_2)^2/c}$ and constants $\sigma_\eta = 0.02$ and $c = 5$. Hence, we assume a spatial correlation of the noise.

As we use a synthetic random process for the noise, we, of course, know the exact solution to the stochastically forced problem. However, in a real world problem we would not be aware of the noise disturbing the system and, hence, would have no knowledge about the actual temperature distribution. In this example, we measure the average temperature $s(z^k) = \frac{1}{|\mathcal{D}_m|} \int_{\mathcal{D}_m} z^k(\boldsymbol{\mu})$ on the small measurement domain $\mathcal{D}_m := \{\boldsymbol{x} \in \mathcal{D} | \|\boldsymbol{x} - (0.5, 0.9)\|_2 \leq 0.025\}$. We assume the measurement data to be subject to synthetic noise, so that the measurement data at time step $k$ is $y_h^k := s(z_h^k) + \vartheta^k$, with $\vartheta$ a Gaussian white noise $\mathbb{R}$-valued stochastic process with variance $R_{\vartheta\vartheta} = 1 \times 10^{-4}$ (that is a standard deviation of about 2% compared to the measurement values). Using this measurement information, we will use the RBKF to obtain state estimates. In particular, we are interested in estimating the average temperature on the domain $\mathcal{D}_1$. Thus, we define the output of interest $y_{int}^k = o(z_h^k) := \int_{\mathcal{D}_1} z^k(\boldsymbol{\mu})$.

To obtain a numerical "true" solution we apply a finite element method on a regular triangular grid with linear ansatz functions spanning the high dimensional space $Z_h$. This results in a discrete model with $H = 1681$ degrees of freedom. Exemplary solutions to the problem without noise for two different parameters are illustrated in Figure 2. The evolution of the not measurable – hence in real world unknown – output of interest is depicted in Figure 1b for an exemplary realization of the stochastically forced model and for comparison an evolution of the output for the same model without noise is plotted. We see, that both outputs deviate remarkably. Hence, there is the need to find a close estimate to the actual solution and the actual output of interest. To that purpose we set up a RBKF. First, we generate a reduced basis space $Z_N$ using the Noise-POD-Greedy Algorithm described in Section 3.1 with $n_{KL} = 5$ KL-modes and with a parameter training set $\mathcal{M}_{train}$ of 25 training points being the vertices of a regular rectangular grid over $\mathcal{P}$. As error estimator we use the energy-norm error bound for the primal solution from (4.26). Having $Z_N$, we obtain the RBKF by projection and can subsequently perform rapid state estimations. An exemplary estimation of the output of interest is depicted in Figure 1b ($red\ line$). As can be seen here, the RBKF provides quite good estimates of the output of interest. Note that we verified the observability of the discretized PDE with the outputs $y_h^k$ $via$ the procedure described in ([12], Thm. 3).

Now, we will analyze this performance in more detail. Because first of all, the RBKF – as well as the ordinary Kalman filter – provides estimates of the noisy solution, we will have a closer look on the performance considering the state estimates before focusing on the output of interest. We will compare the quality of the solution estimates of the RBKF using reduced bases of different sizes to a high dimensional "true" Kalman filter. For this we conducted simulations for a sample of 50 randomly chosen parameters in $\mathcal{P}$. As the system and the RBKF behave in the mean like as if there was no noise, we obtain the expectations by setting $\eta^k = 0$ for all $k = 0, \ldots, K$. For estimating the variance we calculated 50 realizations of noise corrupted simulations

(a)



(b)
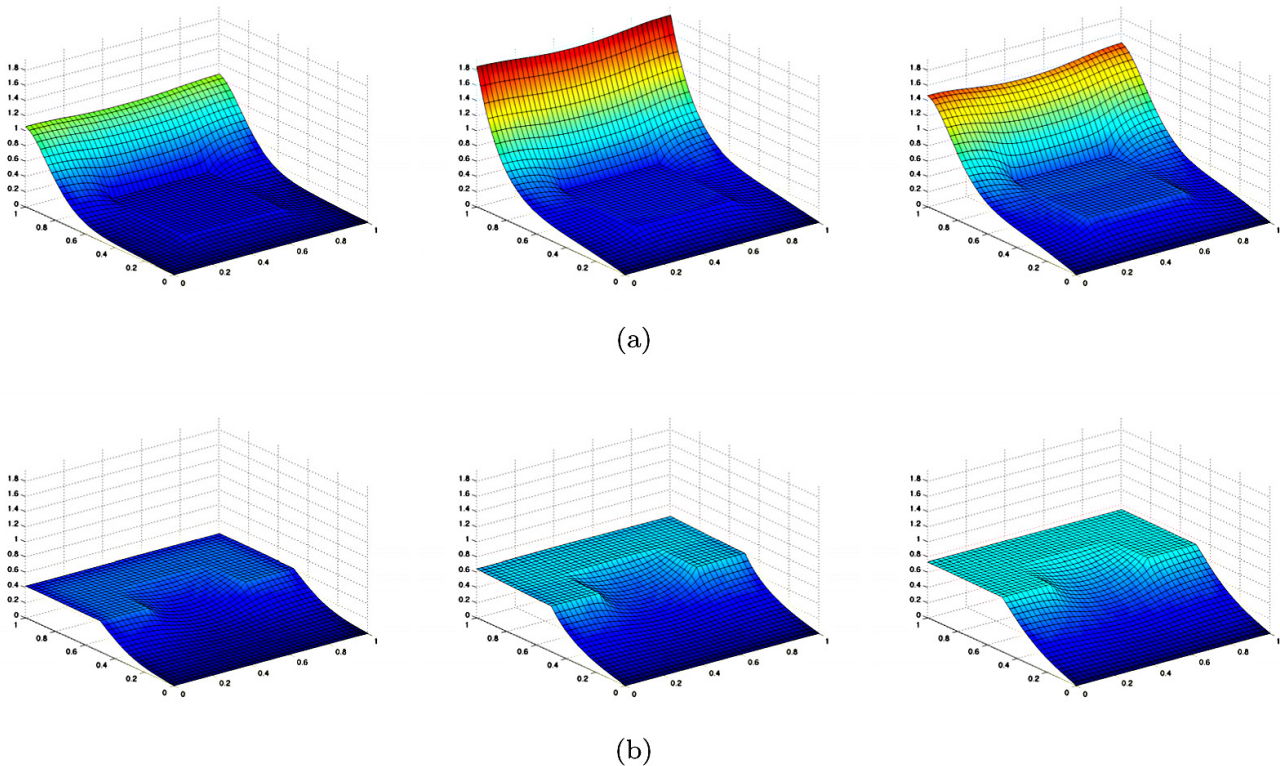
FIGURE 2. Exemplary solutions to the problem without noise disturbances for parameters (a) $\boldsymbol{\mu} = (0.1, 10)^T$ and (b) $\boldsymbol{\mu} = (10, 0.1)^T$ at times $t = 0.5$, $t = 0.8$ and $t = 1$ (from left to right).

for each of the randomly chosen parameters. From the results in Table 1 we see that, the mean relative error of the state estimation error converges to zero for an increasing number $N$ of basis functions. The relative error is below 1% for $N \geq 30$ in the energy norm and the L2-norm. Considering the time needed to perform the state estimation, we see in the third column of Table 1 that the low dimensional RBKF is more than a thousand times faster than the high dimensional Kalman filter, even for a relatively low dimensional "true" model with 1681 degrees of freedom. We see, that evaluating the error bound can be quite expensive in our example, yet, including the rigorous error bound calculation we are still about ten times faster than the full dimensional Kalman filter. However, when using the simplified and not rigorous error estimator described in Section 4.5.1 we are still considerably faster while in practice, we actually do not lose validity of the bound, as we will see later. When applying the method to more complex, and hence higher dimensional problems, the gain in simulation time will be considerably higher, also for the Kalman filter including error estimation.

Next, we analyze the performance of the RBKF focusing on the estimations of the output of interest and the error bounds for the according estimation errors. The results in Table 2 show, that the mean estimation error for the output of interest is below 1% when using 30 basis functions. The same holds for the maximum relative error over the parameter test set. For higher numbers of basis functions $N = N_{\mathrm{du}} = N_{\mathrm{du},o}$ the effectivity $\rho_E = \Delta^k_{\mathrm{Exp}[e_o]} / \mathrm{E}\left[o(z_h^k - \hat{z}_N^k)\right]$ of the error estimator decreases indicating that we obtain a sharp error bound. The results in the table also show, that the bound is rigorous as the minimum effectivity over all parameters and time steps is greater than 1.

While the reduced basis approximation has an influence on the mean of the output estimation error, the variance of this error is not affected by the RB approximation in comparison to the true FEM model, which can be seen in Table 3. This is a desirable property for reduced order filters, as frequently in reduced rank

TABLE 1. In the first two columns the mean over a sample of 50 randomly chosen parameters of the relative expected estimation errors in the energy norm and the $L_2$-norm can be seen. In the last three columns are noted the average computation time needed to perform RBKF or full Kalman filter estimates without the error bound ($t_{KF}$), including computation of the rigorous error bound ($t_{KF\Delta}$) and including the non rigorous error estimator ($t_{KF\Delta,\mathrm{nonrig}}$).

| $N$ | $\mathrm{mean}_{\mu,k}\dfrac{\|\|\mathrm{E}[z_h^k-\hat{z}_N^k]\|\|}{\|\|\mathrm{E}[z_h^k]\|\|}$ | $\mathrm{mean}_{\mu,k}\dfrac{\|\mathrm{E}[z_h^k-\hat{z}_N^k]\|_{L_2}}{\|\mathrm{E}[z_h^k]\|_{L_2}}$ | $t_{KF}[s]$ | $t_{KF\Delta}[s]$ | $t_{KF\Delta,\mathrm{nonrig}}[s]$ |
|---|---|---|---|---|---|
| 10 | 0.0880 | 0.0234 | 0.112 | 1.336 | 0.470 |
| 30 | 0.00758 | 0.00107 | 0.170 | 12.98 | 1.77 |
| 50 | 0.00169 | 0.000216 | 0.258 | 26.07 | 3.80 |
| 70 | 0.000933 | 0.000103 | 0.319 | 38.69 | 5.04 |
| 1681(FEM) | 0 | 0 | 478.4 | – | – |

TABLE 2. Mean relative expected output estimation error and mean and minimum effectivities using basis sizes $N = N_{\mathrm{du}} = N_{\mathrm{du},o}$.

| $N$ | $\mathrm{Mean}_{\mu,k}\dfrac{\mathrm{E}[o(z_h^k-\hat{z}_N^k)]}{\mathrm{E}[o(z_h^k)]}$ | $\mathrm{Maxmean}_{\mu\ \ k}\dfrac{\mathrm{E}[o(z_h^k-\hat{z}_N^k)]}{\mathrm{E}[o(z_h^k)]}$ | $\mathrm{Mean}_{\mu,k}\rho_E$ | $\mathrm{Min}_{\mu,k}\rho_E$ |
|---|---|---|---|---|
| 10 | 0.0233 | 0.105 | 313 | 1.74 |
| 30 | $3.72\times10^{-4}$ | $5.82\times10^{-3}$ | 142 | 1.54 |
| 50 | $1.05\times10^{-4}$ | $2.08\times10^{-3}$ | 45.2 | 1.32 |
| 70 | $3.19\times10^{-5}$ | $5.76\times10^{-4}$ | 6.55 | 1.26 |

TABLE 3. Mean and maximum relative variance of the output estimation error. Furthermore the mean and minimum effectivities with the rigorous and the non rigorous error estimator are indicated.

| $N$ | $\mathrm{Mean}_{\mu,k}\dfrac{\mathrm{Var}(o(z_h^k-\hat{z}_N^k))}{\mathrm{E}[o(z_h^k)]}$ | $\mathrm{Maxmean}_{\mu\ \ k}\dfrac{\mathrm{Var}(o(z_h^k-\hat{z}_N^k))}{\mathrm{E}[o(z_h^k)]}$ | Rigorous $\mathrm{mean}_{\mu,k}\rho_{Var}$ | $\mathrm{min}_{\mu,k}\rho_{Var}$ | Nonrigorous $\mathrm{mean}_{\mu,k}\rho_{Var}$ | $\mathrm{min}_{\mu,k}\rho_{Var}$ |
|---|---|---|---|---|---|---|
| 10 | $1.93\times10^{-3}$ | $1.46\times10^{-2}$ | 4251 | 477 | 4750 | 476 |
| 30 | $1.75\times10^{-3}$ | $0.972\times10^{-2}$ | 2.89 | 1.95 | 2.87 | 1.94 |
| 50 | $1.77\times10^{-3}$ | $0.946\times10^{-2}$ | 1.70 | 1.28 | 1.70 | 1.28 |
| 70 | $1.81\times10^{-3}$ | $0.908\times10^{-2}$ | 1.57 | 1.22 | 1.57 | 1.22 |
| 1681 | $1.82\times10^{-3}$ | $1.41\times10^{-2}$ | – | – | – | – |

approaches divergence can occur, *cf.* [52]. For the variance we again find rapidly computable rigorous error bounds with very good effectivities $\rho_{Var} := \Delta_{\mathrm{Var}(eo)}/\mathrm{Var}(o(z_h^k-\hat{z}_N^k))$, as can be seen in Table 3. Hence, one of the advantages of the RBKF compared to other reduced order Kalman filters is, that there is no divergence in the variance and that this fact can be verified using the rapidly computable bounds. Furthermore, if the reduced order system is observable, the error variance is bounded over time. Although, there is the divergence effect of the RBKF considering the mean estimation error, it can be controlled *via* the error bound.

### 5.1. Experiment considering the basis generation

With a short experiment we will illustrate the advantages of the Noise-POD-Greedy basis generation procedure described in Section 3.1. We generate different reduced bases using the Noise-POD-Greedy algorithm with different numbers of considered KL-modes $n_{KL} \in \{0, 2, 5, 20\}$. Note, that the algorithm with $n_{KL} = 0$ equals the standard POD-Greedy algorithm from [28].
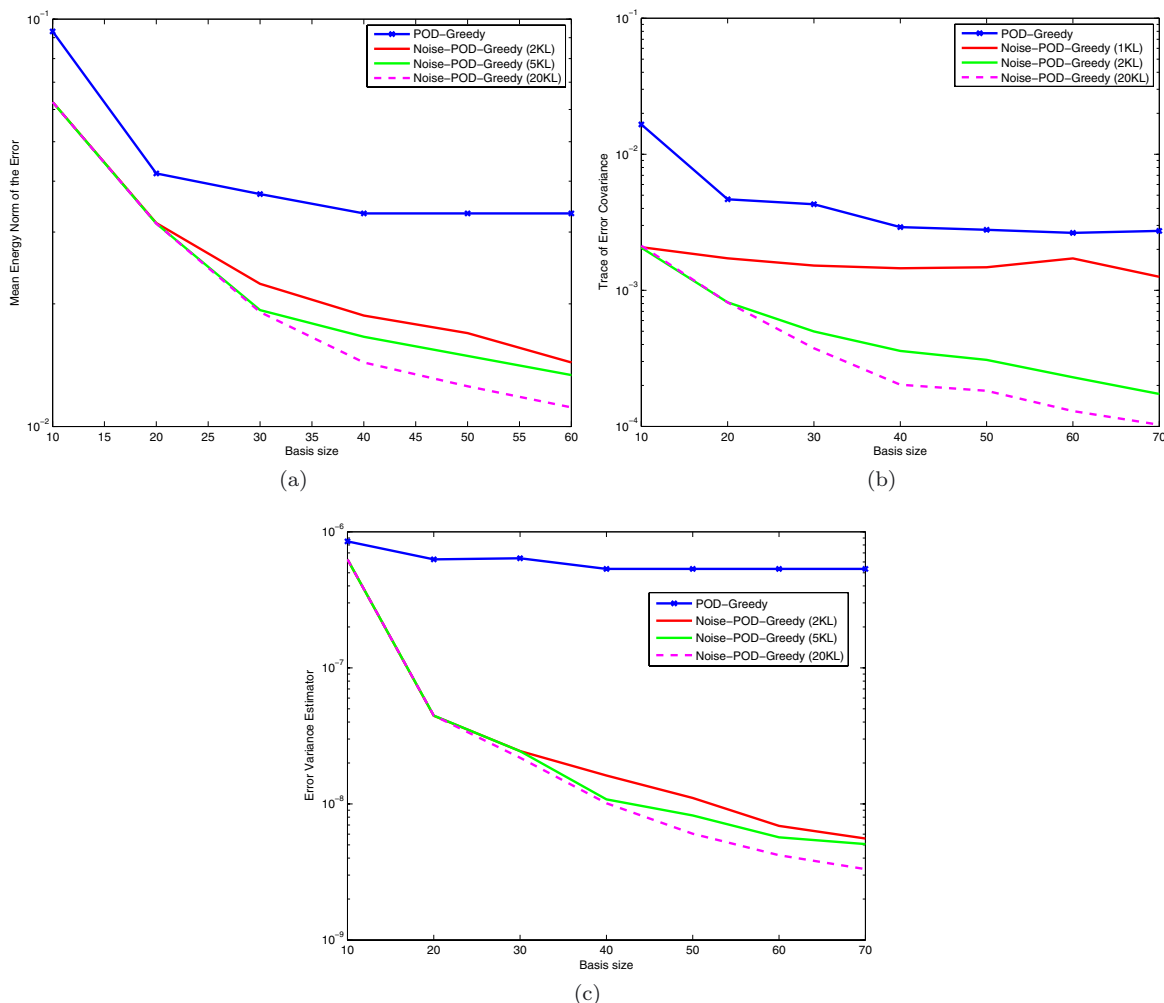
FIGURE 3. For a sample of 50 parameter and 50 noise samples we evaluated (a) the energy norm of the mean error at final time (b) the the trace of the error covariance at final time (c) the variance output error estimator at final time using different basis sizes and for different basis generation procedures using 0, 2, 5 and 20 KL-modes.

Then we investigate on the performance of the respective reduced basis models by running true and reduced order simulations for a sample of 50 randomly chosen test parameters and for each using a sample of 50 noise realizations. Then we investigated on the approximation error by evaluating (a) the energy norm of the mean approximation error, (b) the trace of the error covariance $\text{trace}_{Z_h} \left( (z_h^K - z_N^K) \circ (z_h^K - z_N^K) \right)$ and (c) the error bound $\Delta_{\text{Var}(e_o 1)}^K$ for the output of interest. We see in Figure 3, that when the models are disturbed by a stochastic forcing, the models generated using the Noise-POD-Greedy perform better than the models generated with the standard POD-Greedy. Particularly remarkable is the performance improvement on the variance output error bound. While the standard model stagnates, the error bound decreases with increasing basis size.

Note, as we used a rather high number for the correlation constant $c$ in the design of the noise covariance kernel $\mathcal{K}$, we do not need many KL-modes in order to approximate well the noise characteristics of the disturbances. The number of KL-modes needed in the basis generation process will increase if we lower the spatial correlation of the noise.

## 6. CONCLUSION AND OUTLOOK

We derived the low dimensional reduced basis Kalman filter, providing rapid state estimations for parametrized partial differential equations with stochastic forcing. This low dimensional filter representation is obtained by a projection of the semi-discretized infinite dimensional Kalman filter onto a low dimensional reduced basis space. We presented the new Noise-POD-Greedy algorithm for building this reduced basis space incorporating information about the assumed noise disturbing the system. In order to quantify the quality of the RBKF estimates, we derived error bounds for the expectation and the variance of the estimation error of arbitrary linear output functionals. The stability of the RBKF can be verified *a priori*. In the numerical experiments we demonstrated, that the RBKF performs state estimations considerably faster than the full order Kalman filter, while the loss in accuracy is negligible and can be controlled using the sharp error bounds.

The performance of the RBKF can be improved by applying time adaptive [18] or parameter adaptive [20,27] reduced basis techniques. Thereby, real-time state estimation seems feasible even for complex problems. To take full advantage of the reduced basis method's potential treating parameter dependent problems, one could extend the work presented here to derive parameter adaptive reduced basis extended Kalman filter (*cf*. [10], p. 108*ff*. and [43]) to perform online parameter estimations. Identifying the exact relationship between the here derived time-discrete Kalman filter and the time-continuous Kalman filter for PDEs (*e.g.* [14,15,21]), *i.e.* deriving convergence results, is part of future work.

A weak point of Kalman filtering is, that the characteristics of the noise influencing the system have to be known. To obtain this information noise adaptive strategies [41] can be applied. Another weakness is the potential divergence caused by modeling errors in the FEM model. Applying model reduction to limited memory filters [31] could be an interesting alternative.

Further alternatives to using a low dimensional representation of a Kalman filter could be the application of the reduced basis method to other (time dependent) data assimilation techniques. For example a reduced order 4D-variational formulation for including measurement data [8]. Also accelerating particle filters [47] using reduced basis models could be worth an attempt.

## APPENDIX. PROOFS

### A.1. Proof to Lemma 2.2

*Proof.* We begin with the definition of the Hilbert space $Y^k = \text{span}\{y^1, \ldots, y^k\}$ which is spanned by all outputs at time steps $t^1, \ldots, t^k$ and which has the inner product $(v, w)_{\text{Pr}} := \text{E}[vw]$ for $v, w \in Y^k$. Having a finite dimensional Hilbert space we can apply Gram–Schmidt orthogonalization to find an orthonormal basis $\{\xi^1, \ldots, \xi^k\}$ with

$$\left(\xi^i, \xi^j\right)_{\text{Pr}} = \text{E}\left[\xi^i \xi^j\right] = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{A.1}$$

so that $Y^k = \text{span}\{\xi^1, \ldots, \xi^k\}$. The expression (2.8) of the solution estimation $\hat{z}^k$ in this basis can be reformulated as

$$\hat{z}^k = \sum_{i=1}^{k} a_i \xi^i. \tag{A.2}$$

Here, $a_i \in Z$ and consequently $\hat{z}^k$ is an element of the $k$-dimensional Hilbert space $Y^k(\Omega, \text{Pr}; Z)$. To prove the Lemma we have to show (I) that $\hat{z}^k$ is uniquely determined by (2.10), (II) that this $\hat{z}^k$ is a solution to the filtering problem and (III) that this $\hat{z}^k$ is the unique solution to the filtering problem.

(I) Assume that $\hat{z}^k$ satisfies (2.10). Obviously we can replace the elements $y^j$ from $Y^k$ in (2.10) by $\xi^i$ and reformulate the requirement as

$$\text{E}\left[\left(z^k - \hat{z}^k\right) \xi^i\right] = 0 \tag{A.3}$$

$\forall \xi^i \in Y^k$. Then we obtain from (A.3) by separation, multiplication on both sides with $\xi^i$ and summing up from $i = 1$ to $k$

$$\sum_{i=1}^{k} \mathrm{E}\left[z^k \xi^i\right] \xi^i = \sum_{i=1}^{k} \mathrm{E}\left[\hat{z}^k \xi^i\right] \xi^i. \tag{A.4}$$

As $\mathrm{E}\left[\hat{z}^k \xi^i\right] = \left(\hat{z}^k, \xi^i\right)_{\mathrm{Pr}}$ is the projection of $\hat{z}^k$ onto the basis vector $\xi^i$ it is clear with (A.2) that

$$\sum_{i=1}^{k} \mathrm{E}\left[z^k \xi^i\right] \xi^i = \sum_{i=1}^{k} a_i \xi^i. \tag{A.5}$$

Hence,

$$a_i = \mathrm{E}\left[z^k \xi^i\right] \tag{A.6}$$

meaning the coefficients $a_i$ are the projection of $z^k$ onto the basis vector $\xi^i$ and thereby the estimation $\hat{z}^k$ is uniquely determined and satisfies the constraint (2.8) of the filtering problem.

(II) Next we will show that $\hat{z}^k$ minimizes the expression (2.9) from the filtering problem. For this, we look for a $w \in Y^k(\Omega, \mathrm{Pr}, Z)$ that minimizes

$$\min_{w \in Y^k(\Omega, \mathrm{Pr}; Z)} \left(v, \mathrm{E}\left[(z^k - w) \circ (z^k - w)\right] v\right) \ \forall v \in Z. \tag{A.7}$$

We reformulate

$$\mathrm{E}\left[(z^k - w) \circ \left(z^k - w\right)\right] = \mathrm{E}\left[\left(\tilde{z}^k + \hat{z}^k - w\right) \circ \left(\tilde{z}^k + \hat{z}^k - w\right)\right]$$
$$= \mathrm{E}\left[\tilde{z}^k \circ \tilde{z}^k\right] + \mathrm{E}\left[\left(\hat{z}^k - w\right) \circ \left(\hat{z}^k - w\right)\right] \tag{A.8}$$
$$+ 2 \underbrace{\mathrm{E}\left[\tilde{z}^k \circ \left(\hat{z}^k - w\right)\right]}_{T_1}. \tag{A.9}$$

As $w \in Y^k(\Omega, \mathrm{Pr}; Z)$ we can write $w = \sum_{i=1}^{k} b_i \xi^i$ for some coefficients $b_i \in Z$ and reformulate $T_1$ as

$$T_1 = \mathrm{E}\left[\tilde{z}^k \circ \hat{z}^k\right] - \mathrm{E}\left[\tilde{z}^k \circ w\right]$$
$$= \sum_{i=1}^{k} a_i \mathrm{E}\left[\tilde{z}^k \circ \xi^i\right] - \sum_{i=1}^{k} b_i \mathrm{E}\left[\tilde{z}^k \circ \xi^i\right]. \tag{A.10}$$

From (2.10) and (A.3) we know that $\mathrm{E}\left[\tilde{z}^k \circ \xi\right] = 0 \ \forall \xi \in Y^k$, consequently $T_1 = 0$. Putting this in (A.8) and testing with $v$ we obtain

$$\left(v, \mathrm{E}\left[(z^k - w) \circ (z^k - w)\right] v\right) = \left(v, \mathrm{E}\left[\tilde{z}^k \circ \tilde{z}^k\right] v\right) + \left(v, \mathrm{E}\left[(\hat{z}^k - w) \circ (\hat{z}^k - w)\right] v\right). \tag{A.11}$$

As the first term on the right hand side is independent of the choice of $w$, this expression is minimized for $w = \hat{z}^k$. Hence, $\hat{z}^k$ is a minimizer of (2.9).

(III) It remains to show that $\hat{z}^k$ is the unique minimum of (2.9). Therefore, we first assume there would be another function $\alpha \in Y^k(\Omega, \mathrm{Pr}; Z)$ also minimizing (2.9) so that

$$\left(v, \mathrm{E}\left[(z^k - \alpha) \circ (z^k - \alpha)\right] v\right) \le \left(v, \mathrm{E}\left[(z^k - w) \circ (z^k - w)\right] v\right) \tag{A.12}$$

$\forall v \in Z$ and $\forall w \in Y^k(\Omega, \mathrm{Pr}; Z)$. But we assume that $\alpha$ does not fullfill condition (2.10) and

$$\mathrm{E}\left[(z^k - \alpha)\xi^i\right] = \frac{\beta^i}{n_v} \neq 0 \tag{A.13}$$

with $\beta^i \in Z$ and $n_v \in \mathbb{R}$ a factor to be defined later. In the following the abbreviation $\mathrm{E}\left[(\cdot)^{\circ 2}\right]$ is noted for for $\mathrm{E}\left[(\cdot) \circ (\cdot)\right]$. We can state that

$$\mathrm{E}\left[(z^k - \alpha + \lambda \xi^i)^{\circ 2}\right] = \mathrm{E}\left[(z^k - \alpha)^{\circ 2}\right] + \mathrm{E}\left[(\lambda \xi^i)^{\circ 2}\right] + 2\mathrm{E}\left[(z^k - \alpha) \circ \lambda \xi^i\right]. \tag{A.14}$$

Defining $n_v := \sqrt{(v, \mathrm{E}\left[\xi^{\circ 2}\right] v)} = \sqrt{(v, 1v)} = \|v\|_Z$ and $\lambda := -\frac{\beta^i}{n_v}$ we reformulate (A.14) applied to $v \in Z$ tested with $v \in Z$ while $v \neq 0$:

$$\left(v, \mathrm{E}\left[(z^k - \alpha + \lambda \xi^i)^{\circ 2}\right] v\right) = \left(v, \mathrm{E}\left[(z^k - \alpha)^{\circ 2}\right] v\right) + \left(v, \frac{(\beta^i)^2}{n_v^2} v\right) - 2\left(v, \frac{(\beta^i)^2}{n_v^2} v\right)$$
$$< \left(v, \mathrm{E}\left[(z^k - \alpha)^{\circ 2}\right] v\right). \tag{A.15}$$

Hence, $\alpha$ can not be a minimizer of (2.9) and we proved by contradiction that $\hat{z}^k$ is the unique solution to the filtering problem. $\square$

## A.2. Proof for Proposition 2.4

*Proof.* For the proof that $\hat{z}^k$ from (2.14)–(2.18) is the unique solution to the filtering problem it is sufficient to show that

$$\mathrm{E}\left[\tilde{z}^k y^j\right] = 0 \ \forall j = 1, \ldots, k \tag{A.16}$$

with $\tilde{z}^k = z^k - \hat{z}^k$. This can be concluded from Lemma 2.2. This necessary and sufficient condition (A.16) holds if and only if

$$\left(\mathrm{E}\left[\tilde{z}^k y^j\right], v\right) = 0 \qquad \forall j = 1, \ldots, k \ \forall v \in Z. \tag{A.17}$$

First we have a closer look at the estimation error $\tilde{z}^k$. Subtracting (2.14) from (2.4) and due to linearity of all terms, we obtain the evolution scheme of the estimation error $\tilde{z}$ in weak formulation:

$$m\left(\tilde{z}^k, v\,;\boldsymbol{\mu}\right) + \Delta t\, a_0\left(\tilde{z}^k, v; \boldsymbol{\mu}\,;\boldsymbol{\mu}\right) = m\left(\tilde{z}^{k-1}, v\,;\boldsymbol{\mu}\right) + \left(\eta^k, v\right)$$
$$- l^k(v)\left(y^k - s(\hat{z}_p^k)\right). \tag{A.18}$$

In the following we will often omit the parameter dependence of $m$, $a$ and $\bar{a}$.

Secondly, we will reformulate the expression of the output $y^k$ using the Riesz representation $s(z^k) = \left(s_r, z^k\right)$. Hence with the Riesz representation, (2.7), (2.11) and the fact that $A$ is self adjoint we can reformulate the expression for the output

$$y^k = s(z^k) + \vartheta^k = \left(s_r, z^k\right) + \vartheta^k = \left(AA^{-1}s_r, z^k\right) + \vartheta^k = a\left(z^k, A^{-1}s_r; \boldsymbol{\mu}\right) + \vartheta^k$$
$$\overset{(2.4)}{=} m\left(z^{k-1}, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \Delta t\, b(A^{-1}s_r)u^k + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k. \tag{A.19}$$

In the same way we can state that

$$s(\hat{z}_p^k) = \left(s_r, \hat{z}_p^k\right) = \left(A^{-1}s_r, A\hat{z}_p^k\right) = a\left(\hat{z}_p^k, A^{-1}s_r; \boldsymbol{\mu}\right)$$
$$\overset{(2.15)}{=} m\left(\hat{z}^{k-1}, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \Delta t\, b(A^{-1}s_r)u^k. \tag{A.20}$$

With (A.19) and (A.20) we obtain

$$y^k - s(\hat{z}_p^k) = m\left(\tilde{z}^{k-1}, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k. \tag{A.21}$$

We assume that the state estimation in the preceding time step $\hat{z}^{k-1}$ was also obtained by an orthogonal projection of $z^{k-1}$ onto the approximation space $Y^{k-1} = \mathrm{span}\{y^1, \ldots, y^{k-1}\}$, hence, that $\mathrm{E}\left[\tilde{z}^{k-1} y^j\right] = 0 \ \forall j =$

$1, \ldots, k-1$. So, in the following we will show by induction, that if $\mathrm{E}\left[\tilde{z}^{k-1} y^j\right] = 0 \ \forall j = 1, \ldots, k-1$, then $\mathrm{E}\left[\tilde{z}^k y^j\right] = 0 \ \forall j = 1, \ldots, k$. Afterwards we will show that this same condition is fulfilled for the induction base case $k = 1$. We now reformulate the left hand side of (A.17) to

$$
\begin{aligned}
\left(\mathrm{E}\left[\tilde{z}^k y^j\right], v\right) \ = \ & \left(\mathrm{E}\left[\tilde{z}^k y^j\right], AA^{-1}v\right) = \mathrm{E}\left[\left(A\tilde{z}^k, A^{-1}v\right) y^j\right] = \mathrm{E}\left[a\left(\tilde{z}^k, A^{-1}v; \boldsymbol{\mu}\right) y^j\right] \\
& \stackrel{(\mathrm{A}.18)}{=} \mathrm{E}\left[\left(m\left(\tilde{z}^{k-1}, A^{-1}v \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right) - l^k(A^{-1}v)\left(y^k - s(\hat{z}_p^k)\right)\right)\left(y^j\right)\right] \\
& \stackrel{(\mathrm{A}.21)}{=} \mathrm{E}\bigg[\left(m\left(\tilde{z}^{k-1}, A^{-1}v \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right)\right. \\
& \qquad \left. - l^k(A^{-1}v)\left(m\left(\tilde{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k\right)\right)\left(y^j\right)\bigg]
\end{aligned} \tag{A.22}
$$

$\forall v \in Z$ and for all $j = 1, \ldots, k$. Now by assumption it is $\mathrm{E}\left[\tilde{z}^{k-1} y^j\right] = 0 \ \forall j = 1, \ldots, k-1$ and as $\eta^k$, $\vartheta^k$ and $y^j$ are uncorrelated for all $j = 1, \ldots, k-1$, consequently this expression is zero for all $y^j$ with $j = 1, \ldots, k-1$. Hence, it remains to show that (A.22) is zero for $j = k$. We replace $m\left(z^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) = m\left(\tilde{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + m\left(\hat{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right)$ in (A.19) and obtain with (A.21) and (A.22)

$$
\begin{aligned}
\left(\mathrm{E}\left[\tilde{z}^k y^k\right], v\right) = \ & \mathrm{E}\bigg[\left(m\left(\tilde{z}^{k-1}, A^{-1}v \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right)\right. \\
& \qquad - l^k(A^{-1}v)\left(m\left(\tilde{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k\right)\bigg) \\
& \qquad \times \left(m\left(z^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + \Delta t \, b(A^{-1}s_r)u^k + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k\right)\bigg] \\
= \ & \mathrm{E}\bigg[\left(m\left(\tilde{z}^{k-1}, A^{-1}v \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right)\right. \\
& \qquad - l^k(A^{-1}v)\left(m\left(\tilde{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k\right)\bigg) \\
& \qquad \times \left(m\left(\tilde{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + m\left(\hat{z}^{k-1}, A^{-1}s_r \ ; \boldsymbol{\mu}\right) + \Delta t \, b(A^{-1}s_r)u^k + \left(\eta^k, A^{-1}s_r\right) + \vartheta^k\right)\bigg]
\end{aligned} \tag{A.23}
$$

$\forall v \in Z$. By assumption $\mathrm{E}\left[\tilde{z}^{k-1} \hat{z}^{k-1}\right] = 0$ (due to the fact that $\hat{z}^{k-1} \in Y^{k-1}(\Omega, \mathrm{Pr}; Z)$), knowing that $\mathrm{E}\left[\tilde{z}^k\right] = 0$ (this can easily be seen from (A.18)) and knowing furthermore that the following random functions and random variables are uncorrelated

$$
\begin{aligned}
\mathrm{E}\left[\hat{z}^{k-1} \circ \eta^k\right] = 0 \qquad\qquad & \mathrm{E}\left[\tilde{z}^{k-1} \circ \eta^k\right] = 0 \qquad\qquad && \mathrm{E}\left[\tilde{z}^{k-1} \circ \vartheta^k\right] = 0 \\
\mathrm{E}\left[\eta^k \circ \vartheta^k\right] = 0 \qquad\qquad & \mathrm{E}\left[\hat{z}^{k-1} \circ \vartheta^k\right] = 0 &&
\end{aligned} \tag{A.24}
$$

we can express (A.23) as

$$
\begin{aligned}
\left(\mathrm{E}\left[\tilde{z}^k y^k\right], v\right) = \mathrm{E}\Big[&m\left(\tilde{z}^{k-1}, A^{-1}v\,;\boldsymbol{\mu}\right) m\left(\tilde{z}^{k-1}, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right)\left(\eta^k, A^{-1}s_r\right) \\
&-l^k(A^{-1}v)\left(m\left(\tilde{z}^{k-1}, A^{-1}s_r\,;\boldsymbol{\mu}\right) m\left(\tilde{z}^{k-1}, A^{-1}s_r\,;\boldsymbol{\mu}\right)\right. \\
&\left.+\left(\eta^k, A^{-1}s_r\right)\left(\eta^k, A^{-1}s_r\right) + \vartheta^k\vartheta^k\right)\Big] \\
= \mathrm{E}&\left[\left(MA^{-1}v, \tilde{z}^{k-1}\right)\left(MA^{-1}s_r, \tilde{z}^{k-1}\right)\right] + \mathrm{E}\left[\left(\eta^k, A^{-1}v\right)\left(\eta^k, A^{-1}s_r\right)\right] \\
&-l^k(A^{-1}v)\left(\mathrm{E}\left[\left(MA^{-1}s_r, \tilde{z}^{k-1}\right)\left(MA^{-1}s_r, \tilde{z}^{k-1}\right)\right]\right. \\
&\left.+ \mathrm{E}\left[\left(\eta^k, A^{-1}s_r\right)\left(\eta^k, A^{-1}s_r\right)\right] + \mathrm{E}\left[\vartheta^k\vartheta^k\right]\right) \\
= \left(MA^{-1}v, P^{k-1}MA^{-1}s_r\right) &+ \left(A^{-1}v, R_{\eta\eta}A^{-1}s_r\right) \\
-l^k(A^{-1}v)&\left(\left(MA^{-1}s_r, P^{k-1}MA^{-1}s_r\right) + \left(A^{-1}s_r, R_{\eta\eta}A^{-1}s_r\right) + R_{\vartheta\vartheta}\right)
\end{aligned}
\tag{A.25}
$$

$\forall v \in Z$. If we now plug in the expression for the Kalman gain $l^k(v)$ from (2.16) in (A.25) we obtain

$$
\left(\mathrm{E}\left[\tilde{z}^k y^k\right], v\right) = 0
\tag{A.26}
$$

for all $v \in Z$, proving that $\mathrm{E}\left[\tilde{z}^k y^j\right] = 0 \;\forall j = 1,\ldots,k$ with the choice for $l^k(v)$ from (2.16) and in the case that $\mathrm{E}\left[\tilde{z}^{k-1} y^j\right] = 0 \;\forall j = 1,\ldots,k-1$.

To conclude the proof it remains to show that $\mathrm{E}\left[\tilde{z}^1 y^1\right] = 0$. We can reformulate $\mathrm{E}\left[\tilde{z}^1 y^1\right] = 0$ like in the general case until arriving at equation (A.23). This expression is zero for $k=1$ if $\mathrm{E}\left[\tilde{z}^1 \circ \hat{z}^1\right] = \mathbf{0}$. Hence, we have a closer look at

$$
\begin{aligned}
\left(w, \mathrm{E}\left[\tilde{z}^1 \circ \hat{z}^1\right]v\right) &= \mathrm{E}\left[\left(w, \tilde{z}^1\right)\left(\hat{z}^1, v\right)\right] = \mathrm{E}\left[a\left(A^{-1}w, \tilde{z}^1;\boldsymbol{\mu}\right) a\left(\hat{z}^1, A^{-1}v;\boldsymbol{\mu}\right)\right] \\
&= \mathrm{E}\Big[\left(m\left(\tilde{z}^0, A^{-1}w\,;\boldsymbol{\mu}\right) + \left(\eta^1, A^{-1}w\right)\right. \\
&\quad -l^1(A^{-1}w)\left(m\left(\tilde{z}^0, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \left(\eta^1, A^{-1}s_r\right) + \vartheta^1\right)\big) \\
&\quad \times\left(m\left(\hat{z}^0, A^{-1}v\,;\boldsymbol{\mu}\right) + \Delta t\, b(A^{-1}v;\boldsymbol{\mu})\right. \\
&\quad \left.\left.+ l^1(A^{-1}v)\left(m\left(\tilde{z}^0, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \left(\eta^1, A^{-1}s_r\right) + \vartheta^1\right)\right)\right]
\end{aligned}
\tag{A.27}
$$

where the last equation follows from (2.14) and (A.21). By multiplying out the two inner brackets, many product terms appear. As some random quantities are uncorrelated, *cf.* (A.24), most of them cancel out, leading to

$$
\begin{aligned}
\left(w, \mathrm{E}\left[\tilde{z}^1 \circ \hat{z}^1\right]v\right) = \;&\mathrm{E}\left[m\left(\tilde{z}^0, A^{-1}w\,;\boldsymbol{\mu}\right) m\left(\hat{z}^0, A^{-1}v\,;\boldsymbol{\mu}\right)\right]\left(1 - l^1(A^{-1}w)\right) \\
&+l^1(A^{-1}v)\left(\mathrm{E}\left[m\left(\tilde{z}^0, A^{-1}w\,;\boldsymbol{\mu}\right) m\left(\tilde{z}^0, A^{-1}s_r\,;\boldsymbol{\mu}\right) + \left(\eta^1, A^{-1}w\right)\left(\eta^1, A^{-1}s_r\right)\right]\right) \\
&-l^1(A^{-1}w)l^1(A^{-1}v)\left(\mathrm{E}\Big[m\left(\tilde{z}^0, A^{-1}s_r\,;\boldsymbol{\mu}\right) m\left(\tilde{z}^0, A^{-1}s_r\,;\boldsymbol{\mu}\right)\right. \\
&+\left.\left(\eta^1, A^{-1}s_r\right)\left(\eta^1, A^{-1}s_r\right) + \vartheta^1\vartheta^1\Big]\right)
\end{aligned}
\tag{A.28}
$$

and if we put in for $l^1(A^{-1}w)$ the definition from (2.16) into (A.28), the last three lines of (A.28) become zero. And the term $\mathrm{E}\left[m\left(\tilde{z}^0, A^{-1}w\,;\boldsymbol{\mu}\right) m\left(\hat{z}^0, A^{-1}v\,;\boldsymbol{\mu}\right)\right]$ is zero as the initial estimation $\hat{z}^0 = \mathrm{E}\left[z^0\right]$ is just a constant and $\mathrm{E}\left[\tilde{z}^0\right] = \mathrm{E}\left[z^0 - \hat{z}^0\right] = \mathrm{E}\left[z^0\right] - \mathrm{E}\left[z^0\right] = 0$. Hence, $\mathrm{E}\left[\tilde{z}^1 y^1\right] = 0$ and the proof that this $\hat{z}^k$ is

solution to the filtering problem is almost complete. It remains to determine the derivation of the propagation of the error covariance operator $P^k$. First, we recall the definition of the error covariance operator

$$P^k v = \mathrm{E}\left[\tilde{z}^k\left(\tilde{z}^k, v\right)\right] \ \forall v \in Z. \tag{A.29}$$

Now we can state that

$$
\begin{aligned}
\left(w, P^k v\right) &= \left(w, \mathrm{E}\left[\tilde{z}^k\left(\tilde{z}^k, v\right)\right]\right) = \mathrm{E}\left[\left(w, \tilde{z}^k\right)\left(\tilde{z}^k, v\right)\right] = \mathrm{E}\left[\left(AA^{-1}w, \tilde{z}^k\right)\left(\tilde{z}^k, AA^{-1}v\right)\right] \\
&= \mathrm{E}\left[a\left(\tilde{z}^k, A^{-1}w; \boldsymbol{\mu}\right) a\left(\tilde{z}^k, A^{-1}v; \boldsymbol{\mu}\right)\right] \\
&\stackrel{(A.18)}{=} \mathrm{E}\left[\left(m\left(\tilde{z}^{k-1}, A^{-1}w; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}w\right) - l^k(A^{-1}w)\left(y^k - s(\hat{z}_p^k)\right)\right)\right. \\
&\qquad \left. \times \left(m\left(\tilde{z}^{k-1}, A^{-1}v; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right) - l^k(A^{-1}v)\left(y^k - s(\hat{z}_p^k)\right)\right)\right] \\
&\stackrel{(A.21)}{=} \mathrm{E}\left[m\left(\tilde{z}^{k-1}, A^{-1}w; \boldsymbol{\mu}\right) m\left(\tilde{z}^{k-1}, A^{-1}v; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}w\right)\left(\eta^k, A^{-1}v\right)\right. \\
&\qquad -l^k(A^{-1}w)\left(m\left(\tilde{z}^{k-1}, A^{-1}v; \boldsymbol{\mu}\right) m\left(\tilde{z}^{k-1}, A^{-1}s_r; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}v\right)\left(\eta^k, A^{-1}s_r\right)\right) \\
&\qquad -l^k(A^{-1}v)\left(m\left(\tilde{z}^{k-1}, A^{-1}w; \boldsymbol{\mu}\right) m\left(\tilde{z}^{k-1}, A^{-1}s_r; \boldsymbol{\mu}\right) + \left(\eta^k, A^{-1}w\right)\left(\eta^k, A^{-1}s_r\right)\right) \\
&\qquad +l^k(A^{-1}v)l^k(A^{-1}w)\left(m\left(\tilde{z}^{k-1}, A^{-1}s_r; \boldsymbol{\mu}\right) m\left(\tilde{z}^{k-1}, A^{-1}s_r; \boldsymbol{\mu}\right)\right. \\
&\qquad \left.\left. + \left(\eta^k, A^{-1}s_r\right)\left(\eta^k, A^{-1}s_r\right) + \vartheta^k\vartheta^k\right)\right]
\end{aligned}
$$

$\forall v, w \in Z$. The last three lines of this equation equal zero, which gets obvious if we plug in the expression for $l^k(A^{-1}w)$ from (2.16). This leaves us with

$$
\begin{aligned}
\left(w, P^k v\right) &= \mathrm{E}\left[\left(MA^{-1}w, \tilde{z}^{k-1}\right)\left(\tilde{z}^{k-1}, MA^{-1}v\right) + \left(\eta^k, A^{-1}w\right)\left(\eta^k, A^{-1}v\right)\right. \\
&\qquad \left. -l^k(A^{-1}w)\left(\left(MA^{-1}v, \tilde{z}^{k-1}\right)\left(\tilde{z}^{k-1}, MA^{-1}s_r\right) + \left(\eta^k, A^{-1}v\right)\left(\eta^k, A^{-1}s_r\right)\right)\right] \\
&= \left(MA^{-1}w, P^{k-1}MA^{-1}v\right) + \left(A^{-1}w, R_{\eta\eta}A^{-1}v\right) \\
&\qquad -l^k(A^{-1}w)\left(\left(MA^{-1}s_r, P^{k-1}MA^{-1}v\right) + \left(A^{-1}s_r, R_{\eta\eta}A^{-1}v\right)\right)
\end{aligned} \tag{A.30}
$$

$\forall v, w \in Z$, resulting with (2.17) in the desired error covariance propagation (2.18). $\qquad\square$

## A.3. Proof of Lemma 4.7

*Proof.* First, we show very generally that if $z \in \mathcal{Z}$, then $\mathrm{E}\left[\|z - \mathrm{E}[z]\|^2\right] = \mathrm{trace}_Z\left(\mathrm{cov}(z)\right)$. We define the unbiased random variable $\mathring{z} := z - \mathrm{E}[z]$. Be $\{\psi_i\}_{i=1}^H$ an orthonormal basis for $Z$. With the definition of the trace and the definition of the covariance we obtain

$$
\begin{aligned}
\mathrm{trace}_Z\left(\mathrm{cov}(z)\right) &= \sum_{i=1}^H \left(\psi_i, \mathrm{E}\left[\mathring{z} \circ \mathring{z}\right] \psi_i\right) = \sum_{i=1}^H \left(\psi_i, \mathrm{E}\left[\mathring{z}\left(\mathring{z}, \psi_i\right)\right]\right) \\
&= \mathrm{E}\left[\left(\sum_{i=1}^H \left(\mathring{z}, \psi_i\right)\psi_i, \mathring{z}\right)\right] = \mathrm{E}\left[\left(\mathring{z}, \mathring{z}\right)\right] = \mathrm{E}\left[\|\mathring{z}\|^2\right]
\end{aligned} \tag{A.31}
$$

With (4.25) the covariance operator of the residual can be written as

$$
\begin{aligned}
\left(w, \mathrm{cov}(r^k, r^k)v\right) &= \left(w, \mathrm{E}\left[\mathring{r}^k \circ \mathring{r}^k\right]v\right) \\
&= \left(w, \mathrm{E}\left[(-A\mathring{z}_N^k + M\mathring{z}_N^{k-1} + \eta^k) \circ (-A\mathring{z}_N^k + M\mathring{z}_N^{k-1} + \eta^k)\right]v\right) \\
&= \left(w, \left\{ \underbrace{\mathrm{E}\left[A\mathring{z}_N^k \circ A\mathring{z}_N^k\right]}_{=:T_1} + \underbrace{\mathrm{E}\left[M\mathring{z}_N^{k-1} \circ M\mathring{z}_N^{k-1}\right]}_{=:T_2} + R_{\eta\eta} \right.\right. \\
&\qquad - \underbrace{\mathrm{E}\left[A\mathring{z}_N^k \circ M\mathring{z}_N^{k-1}\right]}_{=:T_3} - \mathrm{E}\left[M\mathring{z}_N^{k-1} \circ A\mathring{z}_N^k\right] - \underbrace{\mathrm{E}\left[A\mathring{z}_N^k \circ \eta^k\right]}_{=:T_4} \\
&\qquad \left.\left. - \mathrm{E}\left[\eta^k \circ A\mathring{z}_N^k\right] + \mathrm{E}\left[M\mathring{z}_N^{k-1} \circ \eta^k\right] + \mathrm{E}\left[\eta^k \circ M\mathring{z}_N^{k-1}\right] \right\}v\right)
\end{aligned}
\tag{A.32}
$$

for all $v, w \in Z_h$.

As $\eta^k$ and $\mathring{z}_N^{k-1}$ are uncorrelated, the covariances $\mathrm{E}\left[M\mathring{z}_N^{k-1} \circ \eta^k\right]$ and $\mathrm{E}\left[\eta^k \circ M\mathring{z}_N^{k-1}\right]$ are zero. Setting $v = w = \psi_i$ and summing up both sides of (A.32) from $i = 1, \ldots, H$ results in the trace of the residual covariance operator

$$
\begin{aligned}
\mathrm{trace}_{Z_h}\left(\mathrm{cov}(r^k, r^k)\right) &= \sum_{i=1}^H \left(\psi_i, \mathrm{cov}(r^k, r^k)\psi_i\right) \\
&= \mathrm{trace}_{Z_h}\left(\mathrm{E}\left[A\mathring{z}_N^k \circ A\mathring{z}_N^k\right]\right) + \mathrm{trace}_{Z_h}\left(\mathrm{E}\left[M\mathring{z}_N^{k-1} \circ M\mathring{z}_N^{k-1}\right]\right) + \mathrm{trace}_{Z_h}\left(R_{\eta\eta}\right) \\
&\quad - 2\mathrm{trace}_{Z_h}\left(\mathrm{E}\left[A\mathring{z}_N^k \circ M\mathring{z}_N^{k-1}\right]\right) - 2\mathrm{trace}_{Z_h}\left(\mathrm{E}\left[A\mathring{z}_N^k \circ \eta^k\right]\right).
\end{aligned}
\tag{A.33}
$$

We reformulate this expression term by term

$$
\begin{aligned}
\mathrm{trace}_{Z_h}\left(\mathrm{E}\left[A\mathring{z}_N^k \circ A\mathring{z}_N^k\right]\right) &= \sum_{l=1}^H \mathrm{E}\left[\sum_{i=1}^N (A\varphi_i, \psi_l)(\mathring{\boldsymbol{z}}_N^k)_i \sum_{j=1}^N (A\varphi_j, \psi_l)(\mathring{\boldsymbol{z}}_N^k)_j\right] \\
&= \sum_{l=1}^H \sum_{i=1}^N \sum_{j=1}^N (A\varphi_i, \psi_l)(A\varphi_j, \psi_l)(\boldsymbol{Q}_N^k)_{i,j} \\
&= \sum_{i=1}^N \sum_{j=1}^N (\boldsymbol{Q}_N^k)_{i,j}\left(A\varphi_i, \sum_{l=1}^H (A\varphi_j, \psi_l)\psi_l\right) \\
&= \sum_{i=1}^N \sum_{j=1}^N (\boldsymbol{Q}_N^k)_{i,j}(\boldsymbol{K}_I)_{j,i} = \sum_{i=1}^N (\boldsymbol{Q}_N^k \boldsymbol{K}_I)_{i,i} \\
&= \mathrm{tr}\left(\boldsymbol{Q}_N^k \boldsymbol{K}_I\right),
\end{aligned}
\tag{A.34}
$$

and analogously $\mathrm{trace}_{Z_h}\left(\mathrm{E}\left[M\mathring{z}_N^{k-1} \circ M\mathring{z}_N^{k-1}\right]\right)$. Furthermore, we find

$$
\begin{aligned}
\mathrm{trace}_{Z_h}\left(\mathrm{E}\left[A\mathring{z}_N^k \circ M\mathring{z}_N^{k-1}\right]\right) &= \sum_{l=1}^H \mathrm{E}\left[\sum_{i=1}^N (A\varphi_i, \psi_l)(\mathring{\boldsymbol{z}}_N^k)_i \sum_{j=1}^N (M\varphi_j, \psi_l)(\mathring{\boldsymbol{z}}_N^{k-1})_j\right] \\
&\stackrel{(4.30)}{=} \sum_{i=1}^N (\boldsymbol{A}_N^{-1}\boldsymbol{M}_N \boldsymbol{Q}_N^{k-1}\boldsymbol{K}_{III})_{i,i} = \mathrm{tr}\left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N \boldsymbol{Q}_N^{k-1}\boldsymbol{K}_{III}\right),
\end{aligned}
\tag{A.35}
$$

$$\text{trace}_{Z_h}\left(\text{E}\left[A\mathring{z}_N^k \circ \eta^k\right]\right) = \sum_{l=1}^{H}\left(\psi_l, \text{E}\left[A\mathring{z}_N^k \circ \eta^k\right]\psi_l\right) = \sum_{l=1}^{H}\text{E}\left[\left(\psi_l, A\mathring{z}_N^k\right)\left(\eta^k, \psi_l\right)\right]$$

$$= \sum_{l=1}^{H}\sum_{i=1}^{N}\text{E}\left[(\mathring{z}_N^k)_i\left(\psi_l, A\varphi_i\right)\left(\eta^k, \psi_l\right)\right]$$

$$\stackrel{(4.30)}{=} \sum_{i=1}^{N}\text{E}\left[\left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\mathring{z}_N^{k-1} + \boldsymbol{A}_N^{-1}\boldsymbol{\eta}_N^k\right)_i\left(\eta^k, A\varphi_i\right)\right]$$

and as $\mathring{z}_N^{k-1}$ and $\eta^k$ are uncorrelated and $(\boldsymbol{\eta}_N^k)_i = \left(\eta^k, \varphi_i\right)$

$$\text{trace}_{Z_h}\left(\text{E}\left[A\mathring{z}_N^k \circ \eta^k\right]\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{A}_N^{-1})_{i,j}\text{E}\left[\left(\eta^k, \varphi_j\right)\left(\eta^k, A\varphi_i\right)\right]$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{A}_N^{-1})_{i,j}\left(\varphi_j, R_{\eta\eta}A\varphi_i\right). \tag{A.36}$$

Putting (A.34)–(A.36) in (A.33) and using (A.31) results in the expression in (4.32).  □

## A.4. Proof to Proposition 4.8

*Proof.* This proof is inspired by ([25], Prop. 4.4). Firstly, we will prove the bound for the measurement output. Therefor we recall the definition of the unbiased solution $\mathring{z}_h^k = z_h^k - \text{E}\left[z_h^k\right]$, the unbiased reduced solution $\mathring{z}_N^k = z_N^k - \text{E}\left[z_N^k\right]$ and the unbiased residual $\mathring{\text{Res}}^k(v) = \text{Res}^k(v) - \text{E}\left[\text{Res}^k(v)\right]$ and find with the definition of the variance

$$\text{Var}\left(y_h^k - y_N^{*,k}\right) = \text{Var}\left(s(z_h^k) - s(z_N^k) - \sum_{i=1}^{k}\text{Res}^i(z_{N,\text{du}}^{K-k+i})\right)$$

$$= \text{E}\Bigg[\Bigg(s(z_h^k) - s(z_N^k) - \sum_{i=1}^{k}\text{Res}^i(z_{N,\text{du}}^{K-k+i})$$

$$- \text{E}\left[s(z_h^k) - s(z_N^k) - \sum_{i=1}^{k}\text{Res}^i(z_{N,\text{du}}^{K-k+i})\right]\Bigg)^2\Bigg]$$

$$= \text{E}\Bigg[\Bigg(s(\mathring{z}_h^k) - s(\mathring{z}_N^k) - \sum_{i=1}^{k}\mathring{\text{Res}}^i(z_{N,\text{du}}^{K-k+i})\Bigg)^2\Bigg]. \tag{A.37}$$

Next, we recall the dual problem for the measurement output

$$m\left(z_{\text{du}}^l, v\,;\boldsymbol{\mu}\right) + \Delta t\,a_0\left(z_{\text{du}}^l, v\,;\boldsymbol{\mu}\right) = m\left(z_{\text{du}}^{l+1}, v\,;\boldsymbol{\mu}\right)$$

using the index $l = 1, \ldots, K$ as time index and we recall the final condition $m\left(z_{\text{du}}^{K+1}, v\,;\boldsymbol{\mu}\right) = s(v)$. Now, we set the time index to $l = K - k + i$ and set $v = \mathring{e}^i := \mathring{z}_h^i - \mathring{z}_N^i$ to obtain

$$m\left(z_{\text{du}}^{K-k+i}, \mathring{e}^i\,;\boldsymbol{\mu}\right) + \Delta t\,a_0\left(z_{\text{du}}^{K-k+i}, \mathring{e}^i\,;\boldsymbol{\mu}\right) = m\left(z_{\text{du}}^{K-k+i+1}, \mathring{e}^i\,;\boldsymbol{\mu}\right)$$

and sum up from $i = 1$ to the actual time step $i = k$

$$\sum_{i=1}^{k}m\left(z_{\text{du}}^{K-k+i}, \mathring{e}^i\,;\boldsymbol{\mu}\right) + \Delta t\,a_0\left(z_{\text{du}}^{K-k+i}, \mathring{e}^i\,;\boldsymbol{\mu}\right) = \sum_{i=1}^{k}m\left(z_{\text{du}}^{K-k+i+1}, \mathring{e}^i\,;\boldsymbol{\mu}\right). \tag{A.38}$$

As we assumed the initial approximation error to be zero $e^0 = 0$, the unbiased initial error $\mathring{e}^0$ is also zero. Then we can rewrite (A.38) by an index shift on the right hand side and using the final condition as

$$\sum_{i=1}^{k} m\left(z_{\text{du}}^{K-k+i}, \mathring{e}^i - \mathring{e}^{i-1} \; ; \boldsymbol{\mu}\right) + \Delta t \, a_0\left(z_{\text{du}}^{K-k+i}, \mathring{e}^i \; ; \boldsymbol{\mu}\right) = m\left(z_{\text{du}}^{K+1}, \mathring{e}^k \; ; \boldsymbol{\mu}\right) = s(\mathring{e}^k). \tag{A.39}$$

In a next step, we have a look at the error evolution of the unbiased error. From (4.22) we derive that

$$m\left(\mathring{e}^i, v \; ; \boldsymbol{\mu}\right) + \Delta t \, a_0\left(\mathring{e}^i, v \; ; \boldsymbol{\mu}\right) = m\left(\mathring{e}^{i-1}, v \; ; \boldsymbol{\mu}\right) + \mathring{\text{Res}}^i(v).$$

If we set $v = z_{\text{du}}^{K-k+i}$ and sum up on both sides from time index $i = 1$ to $k$ we obtain

$$\sum_{i=1}^{k} m\left(\mathring{e}^i - \mathring{e}^{i-1}, z_{\text{du}}^{K-k+i} \; ; \boldsymbol{\mu}\right) + \Delta t \, a_0\left(\mathring{e}^i, z_{\text{du}}^{K-k+i} \; ; \boldsymbol{\mu}\right) = \sum_{i=1}^{k} \mathring{\text{Res}}^i(z_{\text{du}}^{K-k+i}). \tag{A.40}$$

From (A.39) and (A.40) we infer that

$$s(\mathring{e}^k) = \sum_{i=1}^{k} \mathring{\text{Res}}^i(z_{\text{du}}^{K-k+i}), \tag{A.41}$$

hence,

$$s(\mathring{z}_h^k) - s(\mathring{z}_N^k) - \sum_{i=1}^{k} \mathring{\text{Res}}^i(z_{N,\text{du}}^{K-k+i}) = \sum_{i=1}^{k} \mathring{\text{Res}}^i(e_{\text{du}}^{K-k+i}). \tag{A.42}$$

Applying Cauchy–Schwarz and the Hölder's inequality on the right hand side we find

$$s(\mathring{z}_h^k) - s(\mathring{z}_N^k) - \sum_{i=1}^{k} \mathring{\text{Res}}^i(z_{N,\text{du}}^{K-k+i}) \le \left(\sum_{i=1}^{k} \left\|\mathring{r}^i\right\|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{k} \left\|e_{\text{du}}^{K-k+i}\right\|^2\right)^{\frac{1}{2}}. \tag{A.43}$$

We will now reformulate the second term on the right hand side. Using the coercivity of the bilinear form $a_0$ we derive that

$$\sum_{i=1}^{k} \left\|e_{\text{du}}^{K-k+i}\right\|^2 \le \sum_{i=1}^{k} \frac{1}{\bar{\alpha}_0} a_0\left(e_{\text{du}}^{K-k+i}, e_{\text{du}}^{K-k+i} \; ; \boldsymbol{\mu}\right) = \frac{1}{\Delta t \, \bar{\alpha}_0} \sum_{j=K-k+1}^{K} \Delta t \, a_0\left(e_{\text{du}}^j, e_{\text{du}}^j \; ; \boldsymbol{\mu}\right)$$

$$\le \frac{1}{\Delta t \, \bar{\alpha}_0} \sum_{j=K-k+1}^{K} \Delta t \, a_0\left(e_{\text{du}}^j, e_{\text{du}}^j \; ; \boldsymbol{\mu}\right) + m\left(e_{\text{du}}^{K-k+1}, e_{\text{du}}^{K-k+1} \; ; \boldsymbol{\mu}\right)$$

$$= \frac{1}{\Delta t \, \bar{\alpha}_0} |||e_{\text{du}}^{K-k+1}|||_{\text{du}}^2. \tag{A.44}$$

Plugging (A.44) in (A.43) results in

$$s(\mathring{z}_h^k) - s(\mathring{z}_N^k) - \sum_{i=1}^{k} \mathring{\text{Res}}^i(z_{N,\text{du}}^{K-k+i}) \le \left(\frac{1}{\Delta t \, \bar{\alpha}_0} \sum_{i=1}^{k} \left\|\mathring{r}^i\right\|^2\right)^{\frac{1}{2}} \Delta_{\text{du}}^{K-k+1}. \tag{A.45}$$

Finally, using (A.45) in (A.37) and knowing that $\Delta_{\text{du}}^{K-k+1}$ is non-random, leads to the statement in the proposition. The error bound for the output of interest can be derived in the same way using the dual problem of the output of interest. $\qquad\square$

## A.5. Proof to Lemma 4.12

*Proof.* In a first step we reformulate the variance of the sum using the definition of the unbiased residual $\mathring{\mathrm{Res}}^{i}(v)$ from (4.25) and its Riesz representation $\mathring{r}^i$ to obtain

$$
\mathrm{Var}\left(\sum_{i=1}^{k}\mathrm{Res}^{i}(z_{N,\mathrm{du}}^{K-k+i})\right) = \mathrm{E}\left[\left(\sum_{i=1}^{k}\mathrm{Res}^{i}(z_{N,\mathrm{du}}^{K-k+i}) - \mathrm{E}\left[\sum_{i=1}^{k}\mathrm{Res}^{i}(z_{N,\mathrm{du}}^{K-k+i})\right]\right)^2\right]
$$

$$
= \mathrm{E}\left[\left(\sum_{i=1}^{k}\mathring{\mathrm{Res}}^{i}(z_{N,\mathrm{du}}^{K-k+i})\right)^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{k}\left(\mathring{r}^i, z_{N,\mathrm{du}}^{K-k+i}\right)\right)^2\right]
$$

$$
= \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{E}\left[\left(\mathring{r}^i, z_{N,\mathrm{du}}^{K-k+i}\right)\left(\mathring{r}^j, z_{N,\mathrm{du}}^{K-k+j}\right)\right]
$$

$$
= \sum_{i=1}^{k}\sum_{j=1}^{k}(z_{N,\mathrm{du}}^{K-k+i})^T R_{rr}^{i,j} z_{N,\mathrm{du}}^{K-k+j} \tag{A.46}
$$

where in the last line the matrix $R_{rr}^{i,j} \in \mathbb{R}^{N_{\mathrm{du}}\times N_{\mathrm{du}}}$ is defined element-wise as $(R_{rr}^{i,j})_{n,m} := \mathrm{E}\left[\left(\mathring{r}^i, \varphi_{du,m}\right)\left(\mathring{r}^j, \varphi_{du,n}\right)\right]$. Next we show, that $R_{rr}^{i,j}$ can be calculated as stated in (4.47). Therefor we first aim to find an expression for the covariance of the residual, namely $\mathrm{E}\left[\left(\mathring{r}^i, v\right)\left(\mathring{r}^j, w\right)\right]$. We obtain

$$
\mathrm{E}\left[\left(\mathring{r}^i, v\right)\left(\mathring{r}^j, w\right)\right] = \mathrm{E}\left[\left(-a\left(\mathring{z}_N^i, v; \boldsymbol{\mu}\right) + m\left(\mathring{z}_N^{i-1}, v\, ; \boldsymbol{\mu}\right) + \left(\eta^i, v\right)\right)\left(-a\left(\mathring{z}_N^j, w; \boldsymbol{\mu}\right) + m\left(\mathring{z}_N^{j-1}, w\, ; \boldsymbol{\mu}\right) + \left(\eta^j, w\right)\right)\right]
$$

$$
= \mathrm{E}\Bigg[\left(Av, \mathring{z}_N^i\right)\left(\mathring{z}_N^j, Aw\right) + \left(Mv, \mathring{z}_N^{i-1}\right)\left(\mathring{z}_N^{j-1}, Mw\right) + (v, \eta^i)(\eta^j, w)
$$

$$
- \left(Av, \mathring{z}_N^i\right)\left(\mathring{z}_N^{j-1}, Mw\right) - \left(Av, \mathring{z}_N^i\right)(\eta^j, w) - \left(Mv, \mathring{z}_N^{i-1}\right)\left(\mathring{z}_N^j, Aw\right)
$$

$$
+ \left(Mv, \mathring{z}_N^{i-1}\right)(\eta^j, w) - (v, \eta^i)\left(\mathring{z}_N^j, Aw\right) + (v, \eta^i)\left(\mathring{z}_N^{j-1}, Mw\right)\Bigg]
$$

$$
= \left(Av, Q_N^{i,j} Aw\right) + \left(Mv, Q_N^{i-1,j-1} Mw\right) + (v, \mathrm{cov}(\eta^i, \eta^j)w) - \left(Av, Q_N^{i,j-1} Mw\right)
$$

$$
- \left(Mv, Q_N^{i-1,j} Aw\right) - \mathrm{E}\left[\left(Av, \mathring{z}_N^i\right)(\eta^j, w)\right] + \mathrm{E}\left[\left(Mv, \mathring{z}_N^{i-1}\right)(\eta^j, w)\right]
$$

$$
- \mathrm{E}\left[\left(Aw, \mathring{z}_N^j\right)(\eta^i, v)\right] + \mathrm{E}\left[\left(Mv, \mathring{z}_N^{j-1}\right)(\eta^i, w)\right]. \tag{A.47}
$$

with the covariance operator $Q_N^{i,j} : Z \to Z_N$ being defined as $Q_N^{i,j} v := \mathrm{E}\left[\mathring{z}_N^i \left(\mathring{z}_N^j, v\right)\right]$ for all $v \in Z$ and $i, j = 1, \ldots, K$. Note, that $(Q_N^{i,j})_{m,n} = \left(\varphi_m, Q_N^{i,j} \varphi_n\right)$.

Now, we will compare (4.47) and (A.47) term by term. We start with $\left(Av, Q_N^{i,j} Aw\right)$ to find

$$
\left(Av, Q_N^{i,j} Aw\right) = \sum_{p=1}^{N}\sum_{q=1}^{N}(Av, \varphi_p)(Q_N^{i,j})_{p,q}(\varphi_q, Aw). \tag{A.48}
$$

Setting $v = \varphi_{du,n}$ and $w = \varphi_{du,m}$ leads to $\left(A\varphi_{du,m}, Q_N^{i,j} A\varphi_{du,n}\right) = (K_A Q_N^{i,j} (K_A)^T)_{m,n}$ for all $m, n = 1, \ldots, N$. In the same manner $\left(Mv, Q_N^{i-1,j-1} Mw\right)$, $\left(Av, Q_N^{i,j-1} Mw\right)$ and $\left(Mv, Q_N^{i-1,j} Aw\right)$ in (A.47) are reformulated,

leading to

$$\left(M\varphi_{du,m}, Q_N^{i-1,j-1} M\varphi_{du,n}\right) = \left(\boldsymbol{K}_M \boldsymbol{Q}_N^{i-1,j-1} \boldsymbol{K}_M^T\right)_{m,n},$$

$$\left(A\varphi_{du,m}, Q_N^{i,j-1} M\varphi_{du,n}\right) = \left(\boldsymbol{K}_A \boldsymbol{Q}_N^{i,j-1} \boldsymbol{K}_M^T\right)_{m,n}$$

$$\text{and } \left(M\varphi_{du,n}, Q_N^{i-1,j} A\varphi_{du,m}\right) = \left(\boldsymbol{K}_A \boldsymbol{Q}_N^{j,i-1} \boldsymbol{K}_M^T\right)_{m,n}$$

for all $m, n = 1, \ldots, N$. As the white noise is uncorrelated in time, we find $\left(v, \text{cov}(\eta^i, \eta^j)w\right) = \delta_{i,j}\left(v, R_{\eta\eta}w\right)$. Again, setting $v = \varphi_{du,n}$ and $w = \varphi_{du,m}$ leads to the term $\delta_{i,j}\boldsymbol{R}_{\eta\eta\text{du}}$ in the Proposition. Finally, we reformulate

$$\text{E}\left[\left(Av, \mathring{z}_N^i\right)\left(\eta^j, w\right)\right] = \sum_{m=1}^{N} \left(Av, \varphi_m\right) \text{E}\left[(\mathring{\boldsymbol{z}}_N^i)_m \left(\eta^j, w\right)\right] \tag{A.49}$$

using the dof vector $\mathring{\boldsymbol{z}}_N^i$ of the reduced unbiased solution. First, we consider the case $i \geq j$. We recall that if $i \geq j$ then $\mathring{\boldsymbol{z}}_N^i = \left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-j-1}\mathring{\boldsymbol{z}}_N^{j-1} + \sum_{l=j}^{i}\left(\boldsymbol{A}_N^{-1}\boldsymbol{M}_N\right)^{i-l}\boldsymbol{A}_N^{-1}\boldsymbol{\eta}_N^l$ from (4.45). As $\text{cov}(z_N^{j-1}, \eta^j) = 0$ and $\text{cov}(\eta^l, \eta^j) = 0$ for $l \neq j$, putting (4.45) in (A.49) and knowing that $(\boldsymbol{\eta}_N^l)_m = \left(\varphi_m, \eta^l\right)$ leads to

$$\text{E}\left[\left(Av, \mathring{z}_N^i\right)\left(\eta^j, w\right)\right] = \sum_{m=1}^{N}\sum_{n=1}^{N}\left(Av, \varphi_m\right)\left((\boldsymbol{A}_N^{-1}\boldsymbol{M})^{i-j}\boldsymbol{A}_N^{-1}\boldsymbol{R}_{\eta\eta,N}\right)_{m,n}\left(\varphi_n, w\right) \tag{A.50}$$

With $\boldsymbol{R}_{\eta\eta,N}$ defined *via* $(\boldsymbol{R}_{\eta\eta,N})_{m,n} := \left(\varphi_m, R_{\eta\eta}\varphi_n\right)$ as in (3.3). When setting $v = \varphi_{du,p}$ and $w = \varphi_{du,q}$ and varying from $p, q = 1, \ldots, N_{\text{du}}$ this results in $\boldsymbol{K}_A \boldsymbol{R}_{z\eta}^{i,j}$ as stated in the proposition. In case that $i < j$ it is clear from the white noise assumption, that $\text{cov}(z_N^i, \eta^j) = 0$, hence, that $\text{E}\left[\left(Av, \mathring{z}_N^i\right)\left(\eta^j, w\right)\right] = 0$.

Similar arguments can be applied to $\text{E}\left[\left(Mv, \mathring{z}_N^{i-1}\right)\left(\eta^j, w\right)\right]$, $\text{E}\left[\left(Aw, \mathring{z}_N^j\right)\left(\eta^i, v\right)\right]$ and $\text{E}\left[\left(Mv, \mathring{z}_N^{j-1}\right)\left(\eta^i, w\right)\right]$ appearing in (A.47). $\qquad\square$

## A.6. Proof to Proposition 4.14

*Proof.* We recall, that the error $e_{o,3}^k$ is the error between a reduced basis Kalman filter steered by reduced outputs $y_N^k$ and a reduced basis Kalman filter steered by measurement outputs $y_h^k$ for $k = 1, \ldots, K$. It can be written as

$$e_{o,3}^k = o\left(\hat{z}_N^k(y_N) - \hat{z}_N^k(y_h)\right) = \boldsymbol{C}_o^T\left(\hat{\boldsymbol{z}}_N^k(y_N) - \hat{\boldsymbol{z}}_N^k(y_h)\right) = \boldsymbol{C}_o^T \boldsymbol{e}_N^k$$

using the dof-vectors $\hat{\boldsymbol{z}}_N^k(y_N)$, $\hat{\boldsymbol{z}}_N^k(y_h)$ and $\boldsymbol{e}_N^k = \hat{\boldsymbol{z}}_N^k(y_N) - \hat{\boldsymbol{z}}_N^k(y_h)$. Then the error auto-covariance matrix is $\text{cov}(\boldsymbol{e}_N^k) = \text{E}\left[\mathring{\boldsymbol{e}}_N^k(\mathring{\boldsymbol{e}}_N^k)^T\right]$ with the unbiased error $\mathring{\boldsymbol{e}}_N^k = \boldsymbol{e}_N^k - \text{E}\left[\boldsymbol{e}_N^k\right]$. Hence, the variance of the output error can be expressed as

$$\text{Var}\left(e_{o,3}^k\right) = \boldsymbol{C}_o^T\text{E}\left[\mathring{\boldsymbol{e}}_N^k\left(\mathring{\boldsymbol{e}}_N^k\right)^T\right]\boldsymbol{C}_o = \boldsymbol{C}_o^T\boldsymbol{T}\text{E}\left[\boldsymbol{T}^{-1}\mathring{\boldsymbol{e}}_N^k\left(\mathring{\boldsymbol{e}}_N^k\right)^T\boldsymbol{T}^{-T}\right]\boldsymbol{T}^T\boldsymbol{C}_o$$

$$\leq \left|\boldsymbol{C}_o^T\boldsymbol{T}\right|\left|\text{E}\left[\tilde{\mathring{\boldsymbol{e}}}_N^k\left(\tilde{\mathring{\boldsymbol{e}}}_N^k\right)^T\right]\right|\left|\boldsymbol{T}^T\boldsymbol{C}_o\right| \tag{A.51}$$

using the definition $\tilde{\mathring{\boldsymbol{e}}}_N^k := \boldsymbol{T}^{-1}\mathring{\boldsymbol{e}}_N^k$. We recall, that $\boldsymbol{T}$ is the transformation matrix from (4.39). Consequently, to prove the bound stated in (4.48), we have to show that $\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k \geq |\text{E}\left[\tilde{\mathring{\boldsymbol{e}}}_N^k(\tilde{\mathring{\boldsymbol{e}}}_N^k)^T\right]|$ for all $k = 1, \ldots, K$. We will use induction to show, that this statement is true for all time steps.

The base case $k = 0$ is true as $\boldsymbol{P}^0_{\mathring{\boldsymbol{e}}_N} = |\boldsymbol{T}^{-1}\mathrm{E}[\mathring{\boldsymbol{e}}^0_N(\mathring{\boldsymbol{e}}^0_N)^T]\boldsymbol{T}^{-T}| = |\mathrm{E}[\mathring{\tilde{\boldsymbol{e}}}^0_N(\mathring{\tilde{\boldsymbol{e}}}^0_N)^T]|$. In the induction step we show that if $\boldsymbol{P}^{k-1}_{\tilde{\boldsymbol{e}}_N} \geq \mathrm{E}[\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N(\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N)^T]$ then $\boldsymbol{P}^k_{\tilde{\boldsymbol{e}}_N} \geq \mathrm{E}[\mathring{\tilde{\boldsymbol{e}}}^k_N(\mathring{\tilde{\boldsymbol{e}}}^k_N)^T]$. First, we derive the evolution scheme of $\mathring{\boldsymbol{e}}^k_N$ with the evolution scheme of $\boldsymbol{e}^k_N$ from (4.38) and knowing that $\mathring{\boldsymbol{e}}^k_N = \boldsymbol{e}^k_N - \mathrm{E}[\boldsymbol{e}^k_N]$. We then find

$$\mathring{\boldsymbol{e}}^k_N = \boldsymbol{G}_N\mathring{\boldsymbol{e}}^{k-1}_N + \boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\mathring{y}^k - \mathring{y}^{*,k}_N + \sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right)$$

with the unbiased measurement outputs $\mathring{y}^k_h := y^k_h - \mathrm{E}[y^k_h]$, $\mathring{y}^{*,k}_N := y^{*,k}_N - \mathrm{E}[y^{*,k}_N]$ and the unbiased residual defined in (4.25). Hence, the propagation of the error covariance matrix is

$$\mathrm{E}\left[\mathring{\boldsymbol{e}}^k_N\left(\mathring{\boldsymbol{e}}^k_N\right)^T\right] = \boldsymbol{G}_N\mathrm{E}\left[\mathring{\boldsymbol{e}}^{k-1}_N\left(\mathring{\boldsymbol{e}}^{k-1}_N\right)^T\right]\boldsymbol{G}^T_N + \boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T\mathrm{E}\left[\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)^2\right]$$

$$+ \boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T\mathrm{Var}\left(\sum^k_{i=1}\mathrm{Res}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right)$$

$$+ \boldsymbol{G}_N\mathrm{E}\left[\mathring{\boldsymbol{e}}^{k-1}_N\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)\right]\left(\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T + \boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\mathrm{E}\left[\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)\left(\mathring{\boldsymbol{e}}^{k-1}_N\right)^T\right]\boldsymbol{G}^T_N$$

$$+ \boldsymbol{G}_N\mathrm{E}\left[\mathring{\boldsymbol{e}}^{k-1}_N\sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right]\left(\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T$$

$$+ \boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\mathrm{E}\left[\sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\left(\mathring{\boldsymbol{e}}^{k-1}_N\right)^T\right]\boldsymbol{G}^T_N$$

$$+ 2\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T\mathrm{E}\left[\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)\sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right].$$

With the definition of $\mathring{\tilde{\boldsymbol{e}}}^k_N$ this can be transformed to

$$\mathrm{E}\left[\mathring{\tilde{\boldsymbol{e}}}^k_N\left(\mathring{\tilde{\boldsymbol{e}}}^k_N\right)^T\right] = \boldsymbol{D}\mathrm{E}\left[\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N\left(\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N\right)^T\right]\boldsymbol{D}^T + \boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T\mathrm{E}\left[\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)^2\right]$$

$$+ \boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T\mathrm{Var}\left(\sum^k_{i=1}\mathrm{Res}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right)$$

$$+ \boldsymbol{D}\mathrm{E}\left[\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)\right]\left(\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}\boldsymbol{L}^k_N\right)^T$$

$$+ \boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\mathrm{E}\left[\left(\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N\right)^T\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)\right]\boldsymbol{D}^T$$

$$+ \boldsymbol{D}\mathrm{E}\left[\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N\sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right]\left(\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T$$

$$+ \boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\mathrm{E}\left[\left(\mathring{\tilde{\boldsymbol{e}}}^{k-1}_N\right)^T\sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right]\boldsymbol{D}^T$$

$$+ 2\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\left(\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}_N\boldsymbol{L}^k_N\right)^T\mathrm{E}\left[\left(\mathring{y}^k_h - \mathring{y}^{*,k}_N\right)\sum^k_{i=1}\mathring{\mathrm{Res}}^i\left(z^{K-k+i}_{N,\mathrm{du}}\right)\right]$$

Hence,

$$
\begin{aligned}
\mathrm{E}\left[\tilde{\mathring{e}}_N^k\left(\tilde{\mathring{e}}_N^k\right)^T\right] \le{} & |\boldsymbol{D}|\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right|\left|\boldsymbol{D}^T\right|+\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\left|\mathrm{E}\left[\left(\mathring{y}_h^k-\mathring{y}_N^{*,k}\right)^2\right]\right| \\
& +\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\operatorname{Var}\left(\sum_{i=1}^k\mathring{\operatorname{Res}}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)+|\boldsymbol{D}|\underbrace{\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\mathring{y}_h^k-\mathring{y}_N^{*,k}\right)\right]\right|}_{=:T_1}\left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| \\
& +\left|\tilde{\boldsymbol{B}}^k\right|\left|\mathrm{E}\left[\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\left(\mathring{y}^k-\mathring{y}_N^{*,k}\right)\right]\right|\left|\boldsymbol{D}^T\right|+|\boldsymbol{D}|\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\sum_{i=1}^k\mathring{\operatorname{Res}}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right]\right|\left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| \\
& +\left|\tilde{\boldsymbol{B}}^k\right|\left|\mathrm{E}\left[\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\sum_{i=1}^k\mathring{\operatorname{Res}}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right]\right|\left|\boldsymbol{D}^T\right| \\
& +2\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\left|\mathrm{E}\left[\left(\mathring{y}_h^k-\mathring{y}_N^{*,k}\right)\sum_{i=1}^k\mathring{\operatorname{Res}}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right]\right|
\end{aligned}
$$

where we used the definition $\tilde{\boldsymbol{B}}^k:=\boldsymbol{T}^{-1}\boldsymbol{A}^{-1}\boldsymbol{L}_N^k$. The term $T_1$ is a vector in $\mathbb{R}^N$. Using Cauchy–Schwarz it can be bounded element-wise by

$$
(T_1)_i=\left|\mathrm{E}\left[\left(\tilde{\mathring{e}}_N^{k-1}\right)_i\left(\mathring{y}_h^k-\mathring{y}_N^{*,k}\right)\right]\right|\le\sqrt{\mathrm{E}\left[\left(\tilde{\mathring{e}}_N^{k-1}\right)_i^2\right]\mathrm{E}\left[\left(\mathring{y}_h^k-\mathring{y}_N^{*,k}\right)^2\right]}
$$

so that for the whole vector using the bound for the variance of the output (4.33) we have

$$
T_1\le\sqrt{\operatorname{diag}\left(\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right)}\sqrt{\Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu})}.
$$

The other mixed expectation terms can be bounded in the same way, so that

$$
\begin{aligned}
\mathrm{E}\left[\tilde{\mathring{e}}_N^k\left(\tilde{\mathring{e}}_N^k\right)^T\right]\le{} & |\boldsymbol{D}|\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right|\left|\boldsymbol{D}^T\right|+\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu}) \\
& +\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\operatorname{Var}\left(\sum_{i=1}^k\operatorname{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right) \\
& +\sqrt{\Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu})}\Bigg\{|\boldsymbol{D}|\sqrt{\operatorname{diag}\left(\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right|\right)}\left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| \\
& +\left|\tilde{\boldsymbol{B}}^k\right|\sqrt{\operatorname{diag}\left(\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right|\right)}^T\left|\boldsymbol{D}^T\right| \\
& +2\left|\tilde{\boldsymbol{B}}^k\left(\tilde{\boldsymbol{B}}^k\right)^T\right|\sqrt{\operatorname{Var}\left(\sum_{i=1}^k\operatorname{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)}\Bigg\} \\
& +\sqrt{\operatorname{Var}\left(\sum_{i=1}^k\operatorname{Res}^i\left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)}\Bigg\{|\boldsymbol{D}|\sqrt{\operatorname{diag}\left(\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right|\right)}\left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| \\
& +\left|\tilde{\boldsymbol{B}}^k\right|\sqrt{\operatorname{diag}\left(\left|\mathrm{E}\left[\tilde{\mathring{e}}_N^{k-1}\left(\tilde{\mathring{e}}_N^{k-1}\right)^T\right]\right|\right)}^T\left|\boldsymbol{D}^T\right|\Bigg\}.
\end{aligned}
$$

In the induction step we assumed, that $\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1} \geq |\mathrm{E}[\tilde{\tilde{\boldsymbol{e}}}_N^{k-1}(\tilde{\tilde{\boldsymbol{e}}}_N^{k-1})^T]|$, hence, we find

$$
\begin{aligned}
\mathrm{E}\left[\tilde{\tilde{\boldsymbol{e}}}_N^k \left(\tilde{\tilde{\boldsymbol{e}}}_N^k\right)^T\right] \leq & |\boldsymbol{D}| \boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1} \left|\boldsymbol{D}^T\right| + \left|\tilde{\boldsymbol{B}}^k \left(\tilde{\boldsymbol{B}}^k\right)^T\right| \Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu}) \\
& + \left|\tilde{\boldsymbol{B}}^k \left(\tilde{\boldsymbol{B}}^k\right)^T\right| \mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i \left(z_{N,\mathrm{du}}^{K-k+i}\right)\right) \\
& + \sqrt{\Delta_{\mathrm{Var}(y)}^k(\boldsymbol{\mu})} \Bigg\{ |\boldsymbol{D}| \sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)} \left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| \\
& + \left|\tilde{\boldsymbol{B}}^k\right| \sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)^T} |\boldsymbol{D}^T| + 2\left|\tilde{\boldsymbol{B}}^k \left(\tilde{\boldsymbol{B}}^k\right)^T\right| \sqrt{\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i \left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)} \Bigg\} \\
& + \sqrt{\mathrm{Var}\left(\sum_{i=1}^k \mathrm{Res}^i \left(z_{N,\mathrm{du}}^{K-k+i}\right)\right)} \Bigg\{ |\boldsymbol{D}| \sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)} \left|\left(\tilde{\boldsymbol{B}}^k\right)^T\right| \\
& + \left|\tilde{\boldsymbol{B}}^k\right| \sqrt{\mathrm{diag}\left(\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^{k-1}\right)^T} \left|\boldsymbol{D}^T\right| \Bigg\} = \boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k.
\end{aligned}
$$

The result $\mathrm{E}[\tilde{\tilde{\boldsymbol{e}}}_N^k(\tilde{\tilde{\boldsymbol{e}}}_N^k)^T] \leq \boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k$ concludes the induction. We have proven, that $\boldsymbol{P}_{\tilde{\boldsymbol{e}}_N}^k$ is an upper bound for the error covariance matrix $\mathrm{E}[\tilde{\tilde{\boldsymbol{e}}}_N^k(\tilde{\tilde{\boldsymbol{e}}}_N^k)^T]$ for all time steps $k \in \mathbb{K}$ and with (A.51) we then know that (4.48) in the proposition is a valid upper bound. □

## References

[1] A. Aalto, *Spatial discretization error in Kalman filtering for discrete-time infinite dimensional systems*. Technical report (2014).

[2] A.V. Balakrishnan, Stochastic Optimization Theory in Hilbert Spaces – 1. *Appl. Math. Opt.* **1** (1974) 97–120.

[3] P. Binev, A. Cohen, P. Dahmen, R. DeVore, G. Petrova and P. Wojtaszczyk, Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.* **42** (2011) 1457–1472.

[4] S. Boyaval, C. Le Bris, Y. Maday, N.C. Nguyen and A.T. Patera, A reduced basis approach for variational problems with stochastic parameters: Application to heat conduction with variable Robin coefficient. *Comput. Methods Appl. Mech. Engrg.* **198** (2009) 3187–3206.

[5] S. Boyaval, C. Le Bris, T. Lelièvre, Y. Maday, N.C. Nguyen and A.T. Patera, Reduced basis techniques for stochastic problems. *Arch. Comput. Method E* **17** (2010) 435–454.

[6] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer (2011).

[7] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme and G. Turinici, A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: M2AN* **46** (2012) 595–603.

[8] Y. Cao, J. Zhu, I.M. Navon and Z. Luo, A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Meth. Fluids* **53** (2006) 1571–1583.

[9] J. Chandrasekar, I.S. Kim, D.S. Bernstein and A.J. Ridley, Cholesky-Based Reduced-Rank Square-Root Kalman Filtering. In *2008 American Control Conference Seattle, USA* (2008).

[10] C.K. Chui and G. Chen, Kalman Filtering with Real-Time Applications. Springer (2009).

[11] S.E. Cohn, Dynamics of short-term univariate forecast error covariances. *Mon. Weather Rev.* **121** (1993) 3123–3149.

[12] S.E. Cohn and D.P. Dee, Observability of discretized partial differential equations. *SIAM J. Numer. Anal.* **25** (1988) 586–617.

[13] S.E. Cohn and R. Todling, Approximate Data Assimilation Schemes for Stable and Unstable Dynamics. *J. Meteorol. Soc. Jpn* **74** (1996) 63–75.

[14] R. Curtain, A survey of infinite-dimensional filtering. *SIAM Rev.* **17** (1975) 395–411.

[15] R.F. Curtain, Stochastic Evolution Equations with General White Noise Disturbance. *J. Math. Anal. Appl.* **60** (1977) 570–595.

[16] D.N. Daescu and I.M. Navon, A Dual-Weighted Approach to Order Reduction in 4DVAR Data Assimilation. *Mon. Weather Rev.* **136** (2007) 1026–1041.

[17] M. Dihlmann, *Adaptive Reduced Basis Method for Parameter Optimization and State Estimation of Parameterized Evolution Problems*. Ph.D. thesis, University of Stuttgart (2015).

[18] M. Dihlmann, M. Drohmann and B. Haasdonk, Model reduction of parametrized evolution problems using the reduced basis method with adaptive time-partitioning. In *Proc. of ADMOS 2011* (2011).

[19] M. Drohmann and K. Carlberg, *The ROMES method for statistical modeling of reduced-order-model error*. Technical report, Sandia National Labs (2014).

[20] J.L. Eftang, D.J. Knezevic and A.T. Patera, An hp certified reduced basis method for parametrized parabolic partial differential equations. *Math. Comput. Model. Dyn.* **17** (2011) 395–422.

[21] P.L. Falb, Infinite Dimensional Filtering: The Kalman Bucy-Filter in Hilbert Space. *Inform. Control.* **11** (1967) 103–138.

[22] B.F. Farrell and P.J. Ioannou, State estimation using a reduced-order Kalman filter. *J. Atmos. Sci.* **58** (2001) 3666–3680.

[23] R.J. Fitzgerald, Divergence of the Kalman Filter. *IEEE T. Automat. Contr.* **16** (1971) 736–747.

[24] A. Germani, L. Jetto and M. Piccioni, Galerkin approximation of optimal linear filtering of infinite-dimensional linear systems. *SIAM J. Control. Optim.* **26** (1988) 1287–1305.

[25] M.A. Grepl and A.T. Patera, A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: M2AN* **39** (2005) 157–181.

[26] B. Haasdonk, Convergence rates of the POD-Greedy method. *ESAIM: M2AN* **47** (2013) 859–873.

[27] B. Haasdonk, M. Dihlmann and M. Ohlberger, A training set and multiple basis generation approach for parametrized model reduction based on adaptive grids in parameter space. *Math. Comput. Model. Dyn.* **17** (2012) 423–442.

[28] B. Haasdonk and M. Ohlberger, Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: M2AN* **42** (2008) 277–302.

[29] B. Haasdonk, K. Urban and B. Wieland, Reduced Basis Methods for Parameterized Partial Differential Equations with Stochastic Influences Using the Karhunen-Loéve Expansion. *SIAM J. Uncertainty Quantification* **1** (2013) 79–105.

[30] E. Hairer, S.P. Norsett and G. Wanner, Solving Ordinary Differential Equations. *Springer Ser. Comput. Math.* Springer (2009).

[31] A.H. Jazwinski, Stochastic Processes and Filtering Theory. Academic Press (1970).

[32] R.E. Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng-T ASME* **82** (1960) 35–45.

[33] I. Kim, J. Chandrasekar, H.J. Palanthandalam, A. Ridley and D.S. Bernstein. State estimation for large-scale systems based on reduced-order error-covariance propagation. In *Amer. Contr. Conf.* New York (2007).

[34] L.B. Koralov and Y.G. Sinai, Theory of Probability and Random Processes. Springer (2007).

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

[36] W. Lee, D. McDougall and A.M. Stuart, Kalman filtering and smoothing for linear wave equations with model error. *Inverse Probl.* **27** (2011) 095008.

[37] Z. Li and I.M. Navon, Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.* **127** (2001) 661–683.

[38] M. Loève, Probability Theory I. Springer (1987).

[39] Y. Maday and O. Mula, A generalized empirical interpolation method: application of reduced basis techniques to data assimilation. In *Anal. Numer. Partial Differ. Equ.* Springer (2013).

[40] Y. Maday, A.T. Patera, J.D. Penn and M. Yano, A Parametrized-Background Data-Weak approach to variational data assimilation: Formulation, analysis, and application to acoustics. *Int. J. Numer. Meth. Engng.* **102** (2015) 933–965.

[41] R.K. Mehra, On the Identification of Variances and Adaptive Kalman Filtering. *IEEE T. Automat. Contr.* **15** (1970) 175–184.

[42] P. Moireau and D. Chapelle, Reduced-order unscented Kalman filtering with application to parameter identification in large-dimensional systems. *ESAIM: COCV* **17** (2010) 380–405.

[43] P. Moireau, D. Chapelle and P. Le Tallec, Joint state and parameter estimation for distributed mechanical systems. *Comput. Methods. Appl. Mech. Engng.* **197** (2008) 659–677.

[44] M. Morf, G.S. Sighu and T. Kailath, Some New Algortihms for Recursive Estimation in Constant, Linear, Discrete-Time Systems. *IEEE Trans. Automat. Contr.* **19** (1974) 315–323.

[45] K.M. Nagpal, R.E. Helmick and C.S. Sims, Reduced-order estimation Part 1. Filtering. *Int. J. Control* **45** (1987) 1867–1888.

[46] C.F. Price, An Analysis of the Divergence Problem in the Kalman Filter. *IEEE T. Automat. Contr.* **13** (1968) 699–702.

[47] B Ristic, S. Arulampalam and N. Gordon, Beyond the Kalman Filter. Particle Filters for Tracking Applications. Artech House (2004).

[48] E.M. Rønquist and A.T. Patera, Regression on parametric manifolds: Estimation of spatial fields, functional outputs, and parameters from noisy data. *C.R. Acad. Sci. Paris, Ser I* **350** (2012) 543–547.

[49] G. Rozza, D.B.P. Huynh and A.T. Patera, Reduced Basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics. *Arch. Comput. Meth. Eng.* **15** (2008) 229–275.

[50] D. Simon, Optimal State Estimation, Kalman, $H_\infty$, and Nonlinear Approaches. John Wiley&Sons (2006).

[51] H.W. Sorenson, On Applications of Kalman Filtering. *IEEE T. Automat. Contr.* **28** (1983) 254–255.

[52] R. Todling and S.E. Cohn, Suboptimal Schemes for Atmospheric Data Assimilation Based on the Kalman filter. *Monthly Weather Review* **122** (1994) 2530–2557.

[53] K. Urban and A.T. Patera, An improved error bound for reduced basis approximation of linear parabolic problems. *Math. Comput.* **83** (2014) 1599–1615.

[54] M. Verlaan and A.W. Heemink, Tidal flow forecasting using reduced rank square root filters. *Stoch. Hydrol. Hydraul.* **11** (1997) 349–368.

[55] K. Veroy, C. Prud'homme, D.V. Rovas and A.T. Patera, A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In *Proc. of 16th AIAA computational fluid dynamics conference* (2003) 2003–3847.

[56] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press (2010).

[57] K. Yosida, Functional Analysis. Springer (1991).