Journal of Scientific Computing (2023) 97:47
https://doi.org/10.1007/s10915-023-02358-4

Time-Limited Balanced Truncation for Data Assimilation Problems

Josie König[1] · **Melina A. Freitag**[1]

Received: 14 December 2022 / Revised: 25 August 2023 / Accepted: 12 September 2023 /
Published online: 5 October 2023
© The Author(s) 2023

Abstract
Balanced truncation is a well-established model order reduction method which has been applied to a variety of problems. Recently, a connection between linear Gaussian Bayesian inference problems and the system-theoretic concept of balanced truncation has been drawn (Qian et al in Sci Comput 91:29, 2022). Although this connection is new, the application of balanced truncation to data assimilation is not a novel idea: it has already been used in four-dimensional variational data assimilation (4D-Var). This paper discusses the application of balanced truncation to linear Gaussian Bayesian inference, and, in particular, the 4D-Var method, thereby strengthening the link between systems theory and data assimilation further. Similarities between both types of data assimilation problems enable a generalisation of the state-of-the-art approach to the use of arbitrary prior covariances as reachability Gramians. Furthermore, we propose an enhanced approach using time-limited balanced truncation that allows to balance Bayesian inference for unstable systems and in addition improves the numerical results for short observation periods.

Keywords Time-limited balanced truncation · Bayesian inference · 4D-Var · Model reduction

Mathematics Subject Classification 62F15 · 93A15 · 65L99 · 93B11

1 Introduction

One of the greatest challenges of 21st century mathematics is integrating large amounts of data into computational models to obtain novel insights into the state and dynamics of a system. This is the typical framework for data assimilation. Data assimilation combines available, mostly high-dimensional, models with observations at certain time steps, often

✉ Melina A. Freitag
melina.freitag@uni-potsdam.de

Josie König
josie.koenig@uni-potsdam.de

1 Institut für Mathematik, Universität Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

through time-dependent PDEs. These models occur in various domains of application, such as meteorology [18, 25, 26], geosciences [15, 21, 43], and medicine [20, 30]. They are often unstable.

The typical approach to data assimilation problems is Bayesian inference; see, for instance, [17, 35, 53]. We consider the parameters $\boldsymbol{x} \in \mathbb{R}^d$ to be the realisations of a random variable. The distribution of said random variable is what we want to infer. Some *prior* information is given. The time-dependent observations $\boldsymbol{m}$ are related to the parameters via a given forward model operator $\boldsymbol{G}$ with additive noise $\boldsymbol{\epsilon}$:

$$\boldsymbol{m} = \boldsymbol{G}(\boldsymbol{x}) + \boldsymbol{\epsilon}. \tag{1}$$

From (1), the likelihood $\boldsymbol{m}|\boldsymbol{x}$ is derived. Combining this with the measurements, the prior is updated to the posterior distribution of interest $\boldsymbol{x}|\boldsymbol{m}$. Usually, the initial state of a dynamical system, or rather its distribution, is to be inferred.

Often, the state is high-dimensional and therefore forward model simulations are expensive. This problem is well known from systems and control theory, where reduced-order modelling has been established; see, e.g., [1, 6, 7]. Both non-intrusive data-based methods [9, 44] and intrusive model-based methods [2, 28] have been in the spotlight of past and current research. Model-based approaches are promising in the data assimilation context. One such method is balanced truncation (BT) reducing a stable, linear time-invariant system by projecting it onto a subspace of simultaneously easily reachable and observable states. Variants for time-varying systems [34, 50], inhomogeneous systems [3, 29], bilinear systems [4, 19] and even nonlinear systems [31, 32] have been developed. The application of BT to data assimilation has appeared first in the context of 4D-Var [11, 27, 37, 38].

Recently, Qian et al. [46] have reinvigorated interest in the connection between system-theoretic model order reduction and Bayesian inference. Balanced truncation for linear Gaussian Bayesian inference has been considered. In this paper, we relate the results of Qian et al. to the application of BT to 4D-Var. We generalise the existing results on BT for linear Gaussian Bayesian inference to arbitrary Gaussian priors and unstable systems by proposing time-limited balanced truncation (TLBT) for Bayesian inference. The proposed method improves on the state-of-the-art approach, especially for small observation intervals appearing in many data assimilation applications.

The remainder of the paper is structured as follows: At the beginning, the frameworks for two data assimilation problems and the basics of BT are presented; Sect. 2. Sect. 3 introduces how time-limited balanced truncation is applied in these situations. Section 4 presents numerical results before the paper is concluded in Sect. 5 with the discussion of directions for future work.

2 Setting and Background

We consider the linear dynamical system

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}, \tag{2a}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^d$ is the state of the system at time $t \in \mathbb{R}$, and $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ is a linear model operator describing the state evolution. The initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$ is unknown. Noise-polluted outputs $\boldsymbol{m}_k \in \mathbb{R}^{d_{\text{out}}}$ of the system are only available at discrete time points $0 < t_1 < \ldots < t_n$

via the measurement model

$$\boldsymbol{m}_k = \boldsymbol{C}\boldsymbol{x}(t_k) + \boldsymbol{\epsilon}_k = \boldsymbol{C}e^{\boldsymbol{A}t_k}\boldsymbol{x}_0 + \boldsymbol{\epsilon}_k, \tag{2b}$$

with $\boldsymbol{C} \in \mathbb{R}^{d_{\text{out}} \times d}$ the state-to-output operator, typically with $d_{\text{out}} \ll d$. The noise $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_\epsilon)$ is independently and identically distributed additive Gaussian with known positive definite covariance $\boldsymbol{\Gamma}_\epsilon \in \mathbb{R}^{d_{\text{out}} \times d_{\text{out}}}$. The observation interval of the system is $[0, t_n]$. We always start at time 0, with the unknown initial condition, and set the end time to $t_e = t_n$.

The data assimilation task is to infer the unknown initial state from the noisy measurements. To this end, the Bayesian statistical approach is applied and the initial condition is assumed to be a-priori Gaussian distributed:

$$\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{pr}}), \tag{2c}$$

with $\boldsymbol{\Gamma}_{\text{pr}} \in \mathbb{R}^{d \times d}$ a given positive definite prior covariance matrix.

Substituting the Gaussian prior (2c) in the linear Gaussian measurement model (2b) yields a Gaussian likelihood for the measurements conditioned on the initial condition and the measurement times $\boldsymbol{t} = [t_1, \ldots, t_n]^{\text{T}}$:

$$\boldsymbol{m}|(\boldsymbol{x}_0, \boldsymbol{t}) \sim \mathcal{N}(\boldsymbol{G}\boldsymbol{x}_0, \boldsymbol{\Gamma}_{\text{obs}}),$$

where $\boldsymbol{m} \in \mathbb{R}^{nd_{\text{out}}}$, $\boldsymbol{G} \in \mathbb{R}^{nd_{\text{out}} \times d}$, and $\boldsymbol{\Gamma}_{\text{obs}} \in \mathbb{R}^{nd_{\text{out}} \times nd_{\text{out}}}$ are defined as follows:

$$\boldsymbol{m} = \begin{bmatrix} \boldsymbol{m}_1 \\ \vdots \\ \boldsymbol{m}_n \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} \boldsymbol{C}e^{\boldsymbol{A}t_1} \\ \vdots \\ \boldsymbol{C}e^{\boldsymbol{A}t_n} \end{bmatrix}, \quad \boldsymbol{\Gamma}_{\text{obs}} = \begin{bmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}} & & & \\ & \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}} & & \\ & & \ddots & \\ & & & \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}} \end{bmatrix}.$$

For a Gaussian prior and a Gaussian likelihood, the posterior is again Gaussian [14]:

$$\begin{aligned} &\boldsymbol{x}_0|(\boldsymbol{m}, \boldsymbol{t}) \sim \mathcal{N}(\boldsymbol{\mu}_{\text{pos}}, \boldsymbol{\Gamma}_{\text{pos}}), \quad \text{where} \\ &\quad \boldsymbol{\mu}_{\text{pos}} = \boldsymbol{\Gamma}_{\text{pos}}\boldsymbol{G}^{\text{T}}\boldsymbol{\Gamma}_{\text{obs}}^{-1}\boldsymbol{m} \text{ and } \boldsymbol{\Gamma}_{\text{pos}} = (\boldsymbol{H} + \boldsymbol{\Gamma}_{\text{pr}}^{-1})^{-1}. \end{aligned} \tag{3}$$

The matrix $\boldsymbol{H} \in \mathbb{R}^{d \times d}$ is called the Fisher information of $\boldsymbol{m}$ [39] and defined as

$$\boldsymbol{H} = \boldsymbol{G}^{\text{T}}\boldsymbol{\Gamma}_{\text{obs}}^{-1}\boldsymbol{G} = \sum_{k=1}^{n} \boldsymbol{G}_k^{\text{T}}\boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{G}_k = \sum_{k=1}^{n} e^{\boldsymbol{A}^{\text{T}}t_k}\boldsymbol{C}^{\text{T}}\boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{C}e^{\boldsymbol{A}t_k}.$$

The advantage of this data assimilation setting is that the posterior statistics (3) can be obtained in a closed form. The challenge is the computation of the posterior mean and covariance for a high-dimensional state space, that is for d large. This implies several multiplications with the high-dimensional forward map $\boldsymbol{G}$ and its transpose, which might only be available through costly evolution of the underlying dynamical system. Benner, Qiu, and Stoll have proposed a low-rank compression of the posterior covariance [8]. Other than this, the usually small dimension d_{out} of the observations suggests that the dynamics of interest may be modelled accurately when restricted to a low-dimensional subspace. Model order reduction approaches, such as those presented in this paper, aim to reduce the dimensionality of the problem to make computations more efficient while maintaining sufficient accuracy.

2.1 Incremental Four Dimensional Variational Data Assimilation (4D-Var)

4D-Var is a variational data assimilation technique originating in weather forecasting [51]. In 4D-Var, instead of a continuous model evolution, a discrete time-varying system for $k = 0, 1, \ldots, n$ is considered:

$$\boldsymbol{x}_{k+1} = \mathcal{A}_k(\boldsymbol{x}_k), \tag{4a}$$

$$\boldsymbol{m}_k = \mathcal{C}_k(\boldsymbol{x}_k) + \boldsymbol{\epsilon}_k, \tag{4b}$$

where $\boldsymbol{x}_k \in \mathbb{R}^d$ is the state of the system at time $t_k \in \{0, \ldots, t_n\}$, and $\mathcal{A}_k : \mathbb{R}^d \to \mathbb{R}^d$ is the nonlinear model operator that describes the evolution of the state from time t_k to t_{k+1}. The measurement model suffers from centered Gaussian observational noise $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_k)$ for $k = 0, \ldots, n$. The goal is again to infer the initial state $\boldsymbol{x}_0$ from the noisy measurements $\{\boldsymbol{m}_k\}_{k=0,\ldots,n}$. We assume that an a-priori estimate $\boldsymbol{x}_0^b$ of the initial state is given—the so-called *background state*—which is error-prone:

$$\boldsymbol{x}_0 - \boldsymbol{x}_0^b = \boldsymbol{e}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{pr}}). \tag{4c}$$

Linear Gaussian Bayesian inference and 4D-Var are closely related: discretising (2) with equispaced points in time $t_k = k \cdot h$, $k = 0, \ldots, n$ and $t_n = t_e$, we obtain a system of the more general type (4) with $\mathcal{A}_k = \boldsymbol{A}_{\text{disc}} = e^{\boldsymbol{A}h}$ and $\mathcal{C}_k = \boldsymbol{C}$ for all k. A closed form solution to the posterior statistics for the initial state can only be computed for systems of type (4) if $\mathcal{A}_k$ is linear and time-invariant. In 4D-Var, for arbitrary $\mathcal{A}_k$, the maximum a posteriori estimate for $\boldsymbol{x}_0$, i.e., $\boldsymbol{x}_0^* = \arg\min_{\boldsymbol{x}_0} J(\boldsymbol{x}_0)$, is obtained by minimising the cost functional derived from the posterior distribution; see, for instance, [23]:

$$J(\boldsymbol{x}_0) = \frac{1}{2}\left\|\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right\|^2_{\boldsymbol{\Gamma}_{\text{pr}}^{-1}} + \frac{1}{2}\sum_{k=0}^{n} \|\boldsymbol{m}_k - \mathcal{C}_k(\boldsymbol{x}_k)\|^2_{\boldsymbol{R}_k^{-1}} \tag{5a}$$

subject to the forward model dynamics

$$\boldsymbol{x}_{k+1} = \mathcal{A}_k(\boldsymbol{x}_k) \text{ for } k = 0, \ldots, n-1, \tag{5b}$$

where $\boldsymbol{x}_k$ is the state at time step t_k and $\|\boldsymbol{v}\|^2_{\boldsymbol{M}} := \boldsymbol{v}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{v}$. Evaluating the functional $J(\boldsymbol{x}_0)$ is very costly because of the nonlinearity of the operators $\mathcal{A}_k$ and $\mathcal{C}_k$. Incremental 4D-Var overcomes this by using consecutive linearised versions of the operators and minimising a quadratic functional [16]. This approach to minimising (5) is an inexact Gauss-Newton method [36]. Linearisations of $\mathcal{A}_k$ and $\mathcal{C}_k$ about the model state $\boldsymbol{x}_k$ lead to Jacobians $\boldsymbol{A}_k$ and $\boldsymbol{C}_k$ in the so-called *tangent linear model*:

$$\delta\boldsymbol{x}_{k+1} = \boldsymbol{A}_k\delta\boldsymbol{x}_k, \;\; \boldsymbol{d}_k = \boldsymbol{C}_k\delta\boldsymbol{x}_k, \tag{6}$$

where $\boldsymbol{d}_k := \boldsymbol{m}_k - \mathcal{C}_k(\boldsymbol{x}_k)$ for all k and $\delta\boldsymbol{x}_k$ denotes a state perturbation. A discussion of how the observation and prior uncertainties influence (6) follows in Sect. 3.1 The corresponding quadratic cost functional that has to be minimised for $\delta\boldsymbol{x}_0$ is given by

$$\tilde{J}(\delta\boldsymbol{x}_0) = \frac{1}{2}\left\|\delta\boldsymbol{x}_0 - [\boldsymbol{x}_0 - \boldsymbol{x}_0^b]\right\|^2_{\boldsymbol{\Gamma}_{\text{pr}}^{-1}} + \frac{1}{2}\sum_{k=0}^{n} \|\boldsymbol{C}_k\delta\boldsymbol{x}_k - \boldsymbol{d}_k\|^2_{\boldsymbol{R}_k^{-1}} \tag{7a}$$

subject to the forward model dynamics

$$\delta\boldsymbol{x}_{k+1} = \boldsymbol{A}_k\delta\boldsymbol{x}_k \text{ for } k = 0, \ldots, n-1. \tag{7b}$$

Finding the minimum $\delta \boldsymbol{x}_0^*$ of (7) is called *inner loop*. The result of the inner loop is the next increment $\delta \boldsymbol{x}_0^{(i)}$ used to update $\boldsymbol{x}_0^{(i+1)} = \boldsymbol{x}_0^{(i)} + \delta \boldsymbol{x}_0^{(i)}$ in the so-called *outer loop*. The trajectory $\boldsymbol{x}^{(i)} = [\boldsymbol{x}_0^{(i)}, \boldsymbol{x}_1^{(i)}, \ldots, \boldsymbol{x}_n^{(i)}]$ and expected observations are obtained by evaluating the nonlinear model based on the current step's initial condition $\boldsymbol{x}_0^{(i)}$.

Despite the promising idea of obtaining $\boldsymbol{x}_0^*$ incrementally, the computation of (7) is still very costly in high-dimensional settings, even in the linear case. For this reason, the system (6) should be reduced to obtain lower-dimensional matrices replacing $\boldsymbol{A}_k$, $\boldsymbol{C}_k$ and $\boldsymbol{\Gamma}_{\text{pr}}$ in (7). Details on how balanced truncation is used to make the minimisation of this cost functional tractable is discussed in Sect. 3.1

2.2 Model Reduction by Balanced Truncation

Projection-based methods are popular in model order reduction for dynamical systems [6, 7]. The system-theoretic concept of balanced truncation [41, 42] is one of them. We provide some background from systems theory in Sect. 2.2.1 before explaining the proposed method in Sect. 2.2.2 The theoretical explanations in these sections are based on a book by Antoulas [1]. The time-limited approach to balanced truncation is described in detail in Sect. 2.2.3

2.2.1 System-Theoretic Basics

Balanced truncation (BT) is usually considered for a linear time invariant (LTI) system with continuous ($t \in \mathbb{R}_+$) or discrete ($t \in \mathbb{Z}_+$) state equation as follows:

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \quad \text{or} \quad \boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t), \tag{8a}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t), \tag{8b}$$

$$\boldsymbol{x}(0) = \boldsymbol{0}. \tag{8c}$$

The output equation and the initial condition are unified for both cases. Here, $\boldsymbol{x}$, $\boldsymbol{A}$, and $\boldsymbol{C}$ are as in (2). The input of the system $\boldsymbol{u}(t) \in \mathbb{R}^{d_{\text{in}}}$ is through the port $\boldsymbol{B} \in \mathbb{R}^{d \times d_{\text{in}}}$. The observer output $\boldsymbol{y}(t) \in \mathbb{R}^{d_{\text{out}}}$ is without any noise. We, exceptionally, use the notation $\boldsymbol{x}(t+1)$ instead of $\boldsymbol{x}_{k+1}$ for discrete time in this subsection to unify the concept of BT for discrete and continuous time. It is important to notice, however, that the physical meaning of the system matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ differs from continuous time to discrete time: when we discretise the continuous-time system $\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}_{\text{cont}}\boldsymbol{x}(t) + \boldsymbol{B}_{\text{cont}}\boldsymbol{u}(t)$ to $\boldsymbol{x}(t+1) = \boldsymbol{A}_{\text{disc}}\boldsymbol{x}(t) + \boldsymbol{B}_{\text{disc}}\boldsymbol{u}(t)$, it holds that $\boldsymbol{A}_{\text{disc}} = e^{\boldsymbol{A}_{\text{cont}}}$ and $\boldsymbol{B}_{\text{disc}} = \int_0^1 e^{\boldsymbol{A}_{\text{cont}}\tau}\mathrm{d}\tau \boldsymbol{B}_{\text{cont}}$ under the assumption that $\boldsymbol{u}(t)$ is constant on each interval $[k, k+1)$, $k \in \mathbb{N}$.

The idea of BT is to truncate the system by removing the states that are difficult to reach and difficult to observe. The concepts of reachability and observability rely on the so-called reachability and observability Gramians, spanning the (sub)spaces of reachable and observable states. The Gramians are functions of time and the **time-limited reachability Gramian** $\boldsymbol{P}(t) \in \mathbb{R}^{d \times d}$ at time $0 < t < \infty$ is defined by

$$\boldsymbol{P}(t) = \int_0^t e^{\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}e^{\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau \quad \text{for } t \in \mathbb{R}_+ \text{ and} \tag{9a}$$

$$\boldsymbol{P}(t) = \sum_{k=0}^{t-1} \boldsymbol{A}^k\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^k \quad \text{for } t \in \mathbb{Z}_+. \tag{9b}$$

The notion of observability of states at a fixed time t is dual to reachability. The **time-limited observability Gramian** $\boldsymbol{Q}(t) \in \mathbb{R}^{d\times d}$ at time $0 < t < \infty$ is defined by

$$\boldsymbol{Q}(t) = \int_0^t e^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}e^{\boldsymbol{A}\tau}\mathrm{d}\tau \quad \text{for } t \in \mathbb{R}_+ \text{ and} \tag{10a}$$

$$\boldsymbol{Q}(t) = \sum_{k=0}^{t-1}(\boldsymbol{A}^{\mathrm{T}})^k\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}^k \quad \text{for } t \in \mathbb{Z}_+. \tag{10b}$$

The Gramians are associated with the reachability energy and the observability energy of a state, respectively:

$$\|\boldsymbol{x}\|^2_{[\boldsymbol{P}(t)]^{-1}} = \boldsymbol{x}^{\mathrm{T}}[\boldsymbol{P}(t)]^{-1}\boldsymbol{x}, \quad \|\boldsymbol{x}\|^2_{\boldsymbol{Q}(t)} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}(t)\boldsymbol{x}, \tag{11}$$

where $\|\boldsymbol{x}\|^2_{[\boldsymbol{P}(t)]^{-1}}$ is the minimal energy necessary to reach the state $\boldsymbol{x}$ at time t and $\|\boldsymbol{x}\|^2_{\boldsymbol{Q}(t)}$ is the maximal energy produced by the output of an observable state $\boldsymbol{x}$ at time t. The time-limited Gramians are, thus, functions of time informing us how much energy is necessary to reach or observe a state at a fixed, finite time t.

By monitoring the system from initial time to infinity, it is possible to determine how much energy is required to reach or observe a state at *any* point in time. The limits of the finite Gramians (9) and (10) exist only under certain restrictive conditions: the system is required to be *stable*, i.e., the eigenvalues of $\boldsymbol{A}$ lie in the open left half plane for the continuous-time setting or lie inside the unit circle for the discrete-time setting. Then, $e^{\boldsymbol{A}\tau}$ is bounded for $\tau \to \infty$ and, hence, $\boldsymbol{P}(t)$ and $\boldsymbol{Q}(t)$ are bounded for $t \to \infty$ and their limits $\boldsymbol{P}_\infty = \lim_{t\to\infty}\boldsymbol{P}(t)$ and $\boldsymbol{Q}_\infty = \lim_{t\to\infty}\boldsymbol{Q}(t)$ exist; similarly for the discrete-time case. The stable LTI system (8) has **infinite reachability Gramian** $\boldsymbol{P}_\infty \in \mathbb{R}^{d\times d}$ defined by

$$\boldsymbol{P}_\infty = \int_0^\infty e^{\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}e^{\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau \quad \text{for } t \in \mathbb{R}_+ \text{ and} \tag{12a}$$

$$\boldsymbol{P}_\infty = \sum_{k=0}^{\infty}\boldsymbol{A}^k\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^k \quad \text{for } t \in \mathbb{Z}_+. \tag{12b}$$

Besides, it has **infinite observability Gramian** $\boldsymbol{Q}_\infty \in \mathbb{R}^{d\times d}$ defined by

$$\boldsymbol{Q}_\infty = \int_0^\infty e^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}e^{\boldsymbol{A}\tau}\mathrm{d}\tau \quad \text{for } t \in \mathbb{R}_+ \text{ and} \tag{13a}$$

$$\boldsymbol{Q}_\infty = \sum_{k=0}^{\infty}(\boldsymbol{A}^{\mathrm{T}})^k\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}^k \quad \text{for } t \in \mathbb{Z}_+. \tag{13b}$$

The advantage of infinite Gramians is that they can be computed efficiently as the solutions to the continuous-time Lyapunov equations $\boldsymbol{A}\boldsymbol{P}_\infty + \boldsymbol{P}_\infty\boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_\infty + \boldsymbol{Q}_\infty\boldsymbol{A} = -\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$ or to the Lyapunov equations $\boldsymbol{A}\boldsymbol{P}_\infty\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{P}_\infty$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_\infty\boldsymbol{A} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} = \boldsymbol{Q}_\infty$ in discrete time. The drawback is that their use limits the application of BT to stable systems. This is our motivation to discuss an approach to BT for unstable systems, which are common in data assimilation applications.

2.2.2 Balanced Truncation

Projection-based model reduction methods aim to find a suitable low-dimensional subspace of the state space containing the important features. A projection operator projects the

state variables and matrices of the system (8) onto this low-dimensional subspace. In our case, the low-dimensional subspace consists of the states that are both easily reachable and easily observable, quantified by the energies (11). The reachability energy is low and the observability energy is high. The directions obtained by BT, therefore, maximise the generalised Rayleigh quotient $\left(\frac{\|\boldsymbol{x}\|_{\boldsymbol{Q}_\infty}}{\|\boldsymbol{x}\|_{\boldsymbol{P}_\infty^{-1}}}\right)^2$. The dominant eigendirections of the matrix pencil $(\boldsymbol{Q}_\infty, \boldsymbol{P}_\infty^{-1}) =: (\boldsymbol{Q}, \boldsymbol{P}^{-1})$, i.e., the eigendirections $\boldsymbol{v}_k$ satisfying $\boldsymbol{Q}\boldsymbol{v}_k = \delta_k^2 \boldsymbol{P}^{-1}\boldsymbol{v}_k$ and associated with the dominant eigenvalues δ_k^2, maximise the generalised Rayleigh quotient. The eigenvalue square roots $\delta_1 > \delta_2 > \ldots > \delta_d$ are called the *Hankel singular values*.

After having determined the dominant directions $\boldsymbol{v}_k$, the goal is to build a *balancing transformation* $\boldsymbol{T}$ such that $\boldsymbol{T}^{-1}\boldsymbol{P}\boldsymbol{T}^{-\mathrm{T}} = \tilde{\boldsymbol{P}} = \tilde{\boldsymbol{Q}} = \boldsymbol{T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{T}$. If $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{Q}}$, reachability (related to $\tilde{\boldsymbol{P}}$) and observability (related to $\tilde{\boldsymbol{Q}}$) are expressed in a common basis and can be considered simultaneously. The balancing transformation is used as the projector to reduce the system dimensions. Balanced truncation for stable LTI systems is based on the singular value decomposition and given by Algorithm 1.

Algorithm 1 Square Root Balanced Truncation

Input: system matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ of system (8), truncation rank $r \leq d$.
Output: reduced system matrices $\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}$, reduced state $\hat{\boldsymbol{x}}$.

1: Obtain a square root decomposition (e.g., Cholesky) of the Gramians ((12), (13) or others), i.e., $\boldsymbol{P} = \boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}$, $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$, e.g., by low-rank Lyapunov solvers [5].
2: $\boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{Z}^{\mathrm{T}} = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}$ (Singular value decomposition).
3: Truncation $\hat{\boldsymbol{\Delta}} := \mathrm{diag}(\delta_1, \ldots, \delta_r)$, $\hat{\boldsymbol{U}} := \boldsymbol{U}[:, 1:r]$, $\hat{\boldsymbol{Z}} := \boldsymbol{Z}[:, 1:r]$.
4: Balancing transformation $\boldsymbol{T} := \boldsymbol{R}\hat{\boldsymbol{Z}}\hat{\boldsymbol{\Delta}}^{-\frac{1}{2}}$, $\boldsymbol{T}^{-1} := \hat{\boldsymbol{\Delta}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{T}}$.
5: Reduced system $\hat{\boldsymbol{x}} := \boldsymbol{T}^{-1}\boldsymbol{x}$, $\hat{\boldsymbol{A}} := \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$, $\hat{\boldsymbol{B}} := \boldsymbol{T}^{-1}\boldsymbol{B}$, $\hat{\boldsymbol{C}} := \boldsymbol{C}\boldsymbol{T}$.

Remark 2.1 The matrix $\boldsymbol{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_d)$ contains the Hankel singular values and $\boldsymbol{R}\boldsymbol{Z} =: \boldsymbol{V} = \begin{bmatrix}\boldsymbol{v}_1 & \ldots & \boldsymbol{v}_d\end{bmatrix}$ contains the corresponding eigendirections of the pencil $(\boldsymbol{Q}, \boldsymbol{P}^{-1})$.

Remark 2.2 The infinite Gramians $\hat{\boldsymbol{P}}, \hat{\boldsymbol{Q}}$ of the reduced system

$$\frac{\mathrm{d}\hat{\boldsymbol{x}}}{\mathrm{d}t} = \hat{\boldsymbol{A}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{B}}\boldsymbol{u}(t), \quad \hat{\boldsymbol{y}} = \hat{\boldsymbol{C}}\hat{\boldsymbol{x}} \tag{14}$$

are given by $\hat{\boldsymbol{P}} = \hat{\boldsymbol{Q}} = \hat{\boldsymbol{\Delta}}$ and, thus, the system is balanced.

Remark 2.3 The balanced truncation Algorithm 1 can be applied to LTI systems regardless of details, as long as Gramians are defined. The typical applications are stable systems with infinite Gramians where Gramian square roots are obtained by solving Lyapunov equations. In Sect. 3 we discuss adaptions of this algorithm to slightly different Gramians. The next subsection introduces time-limited Gramians for unstable systems.

2.2.3 Time-Limited Balanced Truncation

Time- and frequency-limited balanced truncation has been developed by Gawronski and Juan [24]. This section is based on their work. The idea of time-limited balanced truncation is to use the time-limited Gramians (9) and (10) for a time interval of interest instead of the

infinite Gramians (12) and (13). In the data assimilation problems, we consider the interval $\mathcal{T} = [0, \ldots, t_e = t_n]$ and use the notation $\boldsymbol{P}_{\mathcal{T}} := \boldsymbol{P}(t_e)$ and $\boldsymbol{Q}_{\mathcal{T}} := \boldsymbol{Q}(t_e)$. The time-limited Gramians are the solutions of (modified) Lyapunov equations:

Theorem 2.1 *The time-limited Gramians $\boldsymbol{P}_{\mathcal{T}}$ and $\boldsymbol{Q}_{\mathcal{T}}$ for a continuous LTI system (8) with stable $\boldsymbol{A}$ are the unique, positive-semidefinite solutions to the modified continuous-time Lyapunov equations*

$$\boldsymbol{A}\boldsymbol{P}_{\mathcal{T}} + \boldsymbol{P}_{\mathcal{T}}\boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{B}_{t_e}\boldsymbol{B}_{t_e}^{\mathrm{T}}, \;\; \boldsymbol{B}_{t_e} := e^{\boldsymbol{A}t_e}\boldsymbol{B}, \;\; \textit{and}$$
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\mathcal{T}} + \boldsymbol{Q}_{\mathcal{T}}\boldsymbol{A} = -\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} + \boldsymbol{C}_{t_e}^{\mathrm{T}}\boldsymbol{C}_{t_e}, \;\; \boldsymbol{C}_{t_e} := \boldsymbol{C}e^{\boldsymbol{A}t_e}.$$

The time-limited Gramians $\boldsymbol{P}_{\mathcal{T}}$ and $\boldsymbol{Q}_{\mathcal{T}}$ for a discrete system (8) with stable $\boldsymbol{A}$ are the unique, positive-semidefinite solutions to the modified discrete-time Lyapunov equations

$$\boldsymbol{P}_{\mathcal{T}} = \boldsymbol{A}\boldsymbol{P}_{\mathcal{T}}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}^{t_s}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^{t_s} - \boldsymbol{A}^{t_e}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^{t_e}, \;\; \textit{and}$$
$$\boldsymbol{Q}_{\mathcal{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\mathcal{T}}\boldsymbol{A} + (\boldsymbol{A}^{\mathrm{T}})^{t_s}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}^{t_s} - (\boldsymbol{A}^{\mathrm{T}})^{t_e}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}^{t_e}.$$

The balanced truncation Algorithm 1 is applied to the system (8) using the Gramians $\boldsymbol{P} = \boldsymbol{P}_{\mathcal{T}}$ and $\boldsymbol{Q} = \boldsymbol{Q}_{\mathcal{T}}$. Their matrix square roots may be obtained as the solutions of modified Lyapunov equations from Theorem 2.1. We call this procedure *time-limited balanced truncation (TLBT)*. Using TLBT is arguably the way to go in the setting of data assimilation, since there are only observations in some fixed time-interval $\mathcal{T} = [0, t_e]$. The end time t_e is often comparatively small and we are not interested in the behaviour at an infinite time horizon.

The assumption that the system (8) and hence $\boldsymbol{A}$ is stable is only necessary for the existence of the infinite Gramians $\boldsymbol{P}_\infty$ and $\boldsymbol{Q}_\infty$. The time-limited Gramians $\boldsymbol{P}(t)$ and $\boldsymbol{Q}(t)$ are defined anyways, see (9) and (10). Hence, we drop the stability assumption for TLBT. Redmann and Kürschner [49] have shown that for continuous systems with unstable $\boldsymbol{A}$ and $\Lambda(\boldsymbol{A}) \cap \Lambda(-\boldsymbol{A}) = \emptyset$ the Lyapunov equations from Theorem 2.1 can still be used for the computation of the time-limited Gramians, which is beneficial from a numerical perspective. A drawback is that, even for stable $\boldsymbol{A}$, stability is not preserved for systems (14), which were reduced by TLBT, but additional assumptions or modifications are necessary [24].

In this section, two data assimilation approaches and the basics of balanced truncation and its time-limited version have been explained. In the next Sect. 3, the details of (TL)BT in these two methods are discussed and compared.

3 Time-Limited Balanced Truncation for Data Assimilation

The idea of balanced truncation (BT) from Sect. 2.2 has been applied to 4D-Var for discrete time [11, 27, 38] and to linear Gaussian (LG) Bayesian inference for continuous time [46]. An adaptation of (time-limited) BT to data assimilation is only possible if the underlying dynamical system can be considered as an LTI system for which Gramians are defined. In the following, we describe how this adaptation is done for the inner loop of 4D-Var (Sect. 3.1) and LG Bayesian inference (Sect. 3.2) and introduce time-limited BT for both cases.

3.1 TLBT Within the Inner Loop of Incremental 4D-Var

BT for the 4D-Var method was first proposed by Lawless et al. [37] and then applied to settings with [27] or without [11, 38] model error in the state dynamics (4a). In Sect. 2.1 we

have explained why a reduced order model is needed within the inner loop of incremental 4D-Var.

The system to be truncated is the tangent linear model (6). Let us assume that $\boldsymbol{A}_k = \boldsymbol{A}$ and $\boldsymbol{C}_k = \boldsymbol{C}$ for all k. This is an important assumption and restriction, because BT as given in Sect. 2.2 is only defined for LTI systems. In (6), the noise has been ignored to give the basic idea of incremental 4D-Var. With noise, the tangent model in the inner loop of 4D-Var taking into account the prior is:

$$\delta\boldsymbol{x}_0 = \boldsymbol{e}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{pr}}), \;\; \delta\boldsymbol{x}_{k+1} = \boldsymbol{A}\delta\boldsymbol{x}_k, \tag{15a}$$

$$\boldsymbol{d}_k = \boldsymbol{C}\delta\boldsymbol{x}_k, \tag{15b}$$

making sure that the covariance of the increment $\delta\boldsymbol{x}_0$ corresponds to the prior covariance [27]. Again, we assume the observation error to be time-invariant, i.e., $\boldsymbol{R}_k = \boldsymbol{\Gamma}_\epsilon$ for all k. A reduced version of (15) is:

$$\begin{aligned} \delta\hat{\boldsymbol{x}}_0 = \hat{\boldsymbol{e}}_0 \sim \mathcal{N}(\boldsymbol{0}, \hat{\boldsymbol{\Gamma}}_{\text{pr}}), \;\; \delta\hat{\boldsymbol{x}}_{k+1} &= \hat{\boldsymbol{A}}\delta\hat{\boldsymbol{x}}_k, \\ \hat{\boldsymbol{d}}_k &= \hat{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_k, \end{aligned} \tag{16}$$

where the increment $\delta\boldsymbol{x} \in \mathbb{R}^d$ is projected by $\boldsymbol{T}^{-1}\delta\boldsymbol{x} =: \delta\hat{\boldsymbol{x}} \in \mathbb{R}^r$, with $r \ll d$. The model and observation operators are reduced by $\hat{\boldsymbol{A}} := \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ and $\hat{\boldsymbol{C}} := \boldsymbol{C}\boldsymbol{T}$. The projection also affects the state variables, hence $\hat{\boldsymbol{x}} := \boldsymbol{T}^{-1}\boldsymbol{x}$. This justifies the projected version of the prior covariance matrix $\hat{\boldsymbol{\Gamma}}_{\text{pr}} := \boldsymbol{T}^{-1}\boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{T}^{-\text{T}}$; the observations $\boldsymbol{m}_k$ and the observation error covariance $\boldsymbol{\Gamma}_\epsilon$ remain unchanged. We propose to compute the projection operator $\boldsymbol{T}$ via time-limited balanced truncation.

System (15) is almost of type (8), but the input comes with a covariance and thus a different input port at time $k = 0$. Green [27] proposes to begin the summation for the time-limited Gramians with zero input port at $k = 1$. For time step $k = 0$ the summand $\boldsymbol{A}^0\boldsymbol{\Gamma}_{\text{pr}}(\boldsymbol{A}^{\text{T}})^0 = \boldsymbol{\Gamma}_{\text{pr}}$ is added. With these adjustments, the time-limited Gramians (9b) and (10b) of (15) are:

$$\boldsymbol{P}_{\mathcal{T}}^{\text{4D-Var}} = \boldsymbol{\Gamma}_{\text{pr}} + \sum_{k=1}^{t_e-1} \boldsymbol{A}^k\boldsymbol{0}(\boldsymbol{A}^{\text{T}})^k = \boldsymbol{\Gamma}_{\text{pr}} \text{ and} \tag{17a}$$

$$\boldsymbol{Q}_{\mathcal{T}}^{\text{4D-Var}} = \sum_{k=0}^{t_e-1} (\boldsymbol{A}^{\text{T}})^k\boldsymbol{C}^{\text{T}}\boldsymbol{\Gamma}_\epsilon^{-1}\boldsymbol{C}\boldsymbol{A}^k. \tag{17b}$$

The relevant literature is inconsistent regarding the inclusion of the observation error from (4b) in the observability Gramian for the tangent linear model (15). From comparison to (4b), it holds that $\boldsymbol{d}_k = \boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_\epsilon)$ for the error $\boldsymbol{d}_k := \boldsymbol{m}_k - \mathcal{C}_k(\boldsymbol{x}_k)$. In the theses of Boess and Green [11, 27] the observability Gramian for (15) was given by $\boldsymbol{Q}_\infty^{\text{4D-Var}} = \sum_{k=0}^{\infty}(\boldsymbol{A}^{\text{T}})^k\boldsymbol{C}^{\text{T}}\boldsymbol{\Gamma}_\epsilon\boldsymbol{C}\boldsymbol{A}^k$, which we do not consider correct. Instead, we prefer a time-limited formulation of Lawless et al. [38], which corresponds exactly to (17b). Bernstein et al. [10] state that a model reduction by BT with observability Gramian (17b) minimises the distance error between the ouput $\boldsymbol{d}_k$ of the full tangent linear model (15) and the output $\hat{\boldsymbol{d}}_k$ of the reduced model (16), weighted by the observation precision $\boldsymbol{\Gamma}_\epsilon^{-1}$, i.e., $\lim_{k\to\infty}\mathbb{E}\left((\boldsymbol{d}_k - \hat{\boldsymbol{d}}_k)^{\text{T}}\boldsymbol{\Gamma}_\epsilon^{-1}(\boldsymbol{d}_k - \hat{\boldsymbol{d}}_k)\right)$. Weighting by the inverse of the covariance matrix means "penalizing errors in the approximation [...] more strongly in directions of lower [...] variance." [52, p. 11] This also makes sense here: In directions of low observational covariance, we are fairly confident that our observation is correct. The observation errors $\boldsymbol{d}_k$ of the full 4D-Var system (4) are low and so should their approximations $\hat{\boldsymbol{d}}_k$. If $|\boldsymbol{d}_k - \hat{\boldsymbol{d}}_k|$ are large,

our approximation is very inaccurate and has to be penalised. In contrast, if we are uncertain about the observations, the errors $\boldsymbol{d}_k$ have larger variance. If the approximations $\hat{\boldsymbol{d}}_k$ are relatively inaccurate but within the standard deviation, the reduced system (16) approximates (15) reasonably well.

With this explanation, we will from now on use (17b) as observability Gramian for TLBT in incremental 4D-Var. The time-limited observability Gramian for 4D-Var satisfies a modified discrete Lyapunov equation,

$$\boldsymbol{Q}_{\mathcal{T}}^{\text{4D-Var}} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{Q}_{\mathcal{T}}^{\text{4D-Var}} \boldsymbol{A} + \boldsymbol{C}^{\mathrm{T}} \boldsymbol{\Gamma}_{\epsilon}^{-1} \boldsymbol{C} - (\boldsymbol{A}^{\mathrm{T}})^{t_e} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{\Gamma}_{\epsilon}^{-1} \boldsymbol{C} \boldsymbol{A}^{t_e}. \tag{18}$$

The proof of this equation follows along the same lines as for the deterministic case; see, e.g., [1].

Time-limited balanced truncation is applied to system (15) for the inner loop of incremental 4D-Var with Gramians (17). The balancing transformation $\boldsymbol{T}$ and its inverse $\boldsymbol{T}^{-1}$ from Algorithm 1 are the projection matrices to obtain the reduced system (16). If the linear tangent model (15) does not change during the outer loop iterations, the reduced system (16) may be computed once with TLBT for the entire incremental 4D-Var algorithm, making this technique particularly effective.

Standard BT in data assimilation is limited to stable systems. We overcome this by considering time-limited BT. Another approach for unstable discrete systems is α-bounded balanced truncation, which has been developed for 4D-Var by Boess [12]. The idea of α-bounded BT is to shift the system matrices by a parameter α so that the eigenvalues of $\boldsymbol{A}$ lie inside a disc of radius α around the origin. α-bounded BT requires experimentation on the choice of the parameter α [27] and is only applicable to discrete systems. We have, therefore, focused on a more classical system-theoretic concept and have introduced time-limited BT. In Sect. 4 we demonstrate that this approach additionally improves the accuracy for low end times.

3.2 Balancing Linear Gaussian Bayesian Inference

BT has been extended to continuous linear Gaussian (LG) Bayesian inference by Qian et al. [46]. We generalise the concept to arbitrary prior covariances and unstable system matrices $\boldsymbol{A}$. There are two main differences between the LTI system (8) and the LG Bayesian inference setting (2) that need to be resolved before applying BT:

1. In the LG Bayesian inference setting, there is no input function $\boldsymbol{u}(t)$.
2. Unlike for the LTI system, the observations $\boldsymbol{m}_k$ in Bayesian inference are noisy.

3.2.1 From Compatible to Arbitrary Prior Covariances as Reachability Gramians

Qian et al. [46] have solved the problem of the missing input in Bayesian inference by introducing the notion of *prior-compatibility*. It is not necessary to know the input port $\boldsymbol{B}$ explicitly but only to ensure that there is a $\boldsymbol{B} \in \mathbb{R}^{d \times d_{\text{in}}}$ with $d_{\text{in}} = \operatorname{rank}(\boldsymbol{A}\boldsymbol{\Gamma}_{\text{pr}} + \boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{A}^{\mathrm{T}})$ satisfying $\boldsymbol{A}\boldsymbol{\Gamma}_{\text{pr}} + \boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$. A prior covariance matrix $\boldsymbol{\Gamma}_{\text{pr}}$ fulfilling this is called *prior-compatible*. The prior-compatible $\boldsymbol{\Gamma}_{\text{pr}}$ is used as infinite reachability Gramian for the LTI system with given $\boldsymbol{A}$ and possibly unknown $\boldsymbol{B}$. It is also the time-limited reachability Gramian: $\boldsymbol{P}_{\mathcal{T}}^{\text{LG}} = \boldsymbol{P}_{\infty}^{\text{LG}} = \boldsymbol{\Gamma}_{\text{pr}}$.

Prior-compatibility is not ensured in general. The spectral decomposition of $\boldsymbol{A}\boldsymbol{\Gamma}_{\text{pr}} + \boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{A}^{\mathrm{T}}$ may be used to manipulate the prior covariance and make it prior-compatible [46, Sect. 4.1.2]. In practice, however, the prior distribution is usually not chosen arbitrarily, but based on expert knowledge or accumulated information. Changing the prior to make it compatible and

artificially introducing an input port $\boldsymbol{B}$ may involve a loss of information. We want to justify that any Gaussian prior covariance is a time-limited reachability Gramian.

For the tangent linear model in 4D-Var the time-limited reachability Gramian is given by (17a). By comparison of the discrete linear tangent model (15) and the continuous LG Bayesian inference setting, we can use the continuous version of (17a) to compute the time-limited reachability Gramian for (2), which is

$$\boldsymbol{P}_{\mathcal{T}}^{\mathrm{LG}} = \int_0^{t_e} e^{\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}e^{\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau + \boldsymbol{\Gamma}_{\mathrm{pr}} = \int_0^{t_e} e^{\boldsymbol{A}\tau}\boldsymbol{0}\boldsymbol{0}^{\mathrm{T}}e^{\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau + \boldsymbol{\Gamma}_{\mathrm{pr}} = \boldsymbol{\Gamma}_{\mathrm{pr}}. \tag{19a}$$

The above comparison with established results from 4D-Var (see Sect. 3.1) overcomes the need for prior-compatibility for $\boldsymbol{\Gamma}_{\mathrm{pr}}$ and makes the application of BT to LG Bayesian inference more general. A numerical example to emphasise that the original prior should be used instead of a modified compatible one whenever possible is given in Sect. 4.1

This approach is also well-founded in literature: time-limited reachability Gramians are the covariance matrices of the LTI system state at a fixed time [48] and the infinite Gramian is, thus, the total state covariance. The state Eq. (2a) evolves linearly without model error. Hence, the state evolution is deterministic and the full state covariance is introduced by the prior covariance as reachability Gramian. The POD-Gramian of an input-free system is constructed similarly [1].

Remark 3.1 The unknown initial condition to be inferred is not zero. It has a centered Gaussian prior and a Gaussian posterior distribution. BT is defined for a zero initial condition. It may therefore be questioned whether this method is well suited for application to data assimilation problems. Other systems theory techniques, however, require an input matrix; see, e.g., [1, 6, 7]. This input would have to be artificially constructed by modifying prior knowledge, which is not necessary for BT. Therefore, we consider balanced truncation to be the best method so far.

3.2.2 A Noisy Time-Limited Observability Gramian

Regarding the pollution of the measurement $\boldsymbol{m}_k$ by noise $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_\epsilon)$ in the LG Bayesian inference setting (2) compared to the LTI system (8), the following is proposed [46, p. 28]: Recall that $\|\boldsymbol{x}\|_{\boldsymbol{Q}}^2 = \|\boldsymbol{y}\|_{L^2(\mathbb{R})}^2$ defines the maximal energy produced by an observable state $\boldsymbol{x}$. It is the squared distance of the output signal $\boldsymbol{y}$ to $\boldsymbol{0}$ in the $L^2(\mathbb{R})$-norm. For noisy observations, instead of $\|\cdot\|_{L^2(\mathbb{R})}^2$, the Mahalanobis distance [40] from the equilibrium state $\boldsymbol{0}$ to the conditional distribution of a single measurement $\boldsymbol{m}_k|(\boldsymbol{x}_0, \boldsymbol{t}_k) \sim \mathcal{N}(\boldsymbol{C}e^{\boldsymbol{A}t_k}\boldsymbol{x}_0, \boldsymbol{\Gamma}_\epsilon)$ is used, i.e.,

$$d\left(\boldsymbol{0}, \mathcal{N}(\boldsymbol{C}e^{\boldsymbol{A}t_k}\boldsymbol{x}_0, \boldsymbol{\Gamma}_\epsilon)\right) = \left\|\boldsymbol{C}e^{\boldsymbol{A}t_k}\boldsymbol{x}_0\right\|_{\boldsymbol{\Gamma}_\epsilon^{-1}}.$$

We use the same reasoning for the definition of the **noisy time-limited observability Gramian** $\boldsymbol{Q}_{\mathcal{T}}^{\mathrm{LG}} \in \mathbb{R}^{d\times d}$ for linear Gaussian Bayesian inference as:

$$\boldsymbol{Q}_{\mathcal{T}}^{\mathrm{LG}} = \int_0^{t_e} e^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Gamma}_\epsilon^{-1}\boldsymbol{C}e^{\boldsymbol{A}\tau}\mathrm{d}\tau. \tag{19b}$$

It is the unique, positive-definite solution to the dual modified Lyapunov equation

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\mathcal{T}}^{\mathrm{LG}} + \boldsymbol{Q}_{\mathcal{T}}^{\mathrm{LG}}\boldsymbol{A} = -\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Gamma}_\epsilon^{-1}\boldsymbol{C} + \boldsymbol{C}_{t_e}^{\mathrm{T}}\boldsymbol{\Gamma}_\epsilon^{-1}\boldsymbol{C}_{t_e},\ \ \boldsymbol{C}_{t_e} := \boldsymbol{C}e^{\boldsymbol{A}t_e}, \tag{20}$$

similar to the discrete Lyapunov equation (18) for 4D-Var. Remark that $\boldsymbol{Q}_{\mathcal{T}}^{\text{4D-Var}}$ is also a *noisy* (discrete) time-limited observability Gramian. The Fisher matrix $\boldsymbol{H} = \sum_{k=1}^{n} e^{\boldsymbol{A}^{\mathrm{T}} t_k} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{C} e^{\boldsymbol{A} t_k}$ is a discrete summation approach to this same Gramian and could be used as such [46]. This resembles an optimal low-rank method for LG Bayesian inference [52] not based on systems theory.

For (TL)BT, Algorithm 1 is applied to the LG Bayesian inference problem by using the Gramians (19) in Step 1. This leads to the reduced system

$$\begin{aligned} \frac{d\hat{\boldsymbol{x}}}{dt} &= \hat{\boldsymbol{A}}\hat{\boldsymbol{x}}, \text{ especially } \hat{\boldsymbol{x}}(0) = \boldsymbol{T}^{-1}\boldsymbol{x}_0, \\ \boldsymbol{m}_k &\approx \hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \boldsymbol{\epsilon}_k, \ \ \boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_\epsilon) \text{ as before }, \end{aligned} \tag{21}$$

with reduced forward map $\boldsymbol{G}_{(\text{TL})\text{BT}}$ and approximate Fisher information $\boldsymbol{H}_{(\text{TL})\text{BT}}$:

$$\boldsymbol{G}_{(\text{TL})\text{BT}} = \begin{bmatrix} \hat{\boldsymbol{C}} e^{\hat{\boldsymbol{A}} t_1} \boldsymbol{T}^{-1} \\ \vdots \\ \hat{\boldsymbol{C}} e^{\hat{\boldsymbol{A}} t_n} \boldsymbol{T}^{-1} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{C}} e^{\hat{\boldsymbol{A}} t_1} \\ \vdots \\ \hat{\boldsymbol{C}} e^{\hat{\boldsymbol{A}} t_n} \end{bmatrix} \boldsymbol{T}^{-1}, \ \ \boldsymbol{H}_{(\text{TL})\text{BT}} = \boldsymbol{G}_{(\text{TL})\text{BT}}^{\mathrm{T}} \boldsymbol{\Gamma}_{\text{obs}}^{-1} \boldsymbol{G}_{(\text{TL})\text{BT}},$$

and gives the posterior mean and covariance approximations:

$$\begin{aligned} \boldsymbol{\mu}_{\text{pos},(\text{TL})\text{BT}} &= \boldsymbol{\Gamma}_{\text{pos},(\text{TL})\text{BT}} \boldsymbol{G}_{(\text{TL})\text{BT}}^{\mathrm{T}} \boldsymbol{\Gamma}_{\text{obs}}^{-1} \boldsymbol{m}, \\ \boldsymbol{\Gamma}_{\text{pos},(\text{TL})\text{BT}} &= (\boldsymbol{H}_{(\text{TL})\text{BT}} + \boldsymbol{\Gamma}_{\text{pr}}^{-1})^{-1}. \end{aligned} \tag{22}$$

The notation $(\boldsymbol{\mu}_{\text{pos,TLBT}}, \boldsymbol{\Gamma}_{\text{pos,TLBT}})$ denotes a reduced posterior obtained by TLBT and $(\boldsymbol{\mu}_{\text{pos,BT}}, \boldsymbol{\Gamma}_{\text{pos,BT}})$ a reduced posterior obtained by BT with infinite Gramians.

Remark 3.2 For initial state $\boldsymbol{x}(0) = \boldsymbol{0}$ there exists an error bound for the output of the BT-reduced system (21) compared to the full system [1]. For TLBT, error bounds have been derived for continuous time [47, 49] and discrete time [45]. These results are hard to adapt to Bayesian inference, since we infer an unknown non-zero initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$.

Remark 3.3 BT for inhomogeneous, i.e., non-zero, initial conditions has been investigated by Heinkenschloss et al. proposing to augment the input matrix of the LTI system (8) by the initial condition before BT is applied [29]. This technique seems to be less useful for model reduction with changing or unknown initial conditions. Beattie et al. use splitting: the output is approximated by the simultaneous reduction of two systems, one depending on the initial state $\boldsymbol{x}_0 \neq \boldsymbol{0}$ and one with homogeneous initial condition [3]. Regarding the application to Bayesian inference, we have a different objective: not the output approximation, but the reduced posterior statistics.

3.2.3 Computational Aspects of TLBT in Data Assimilation

The use of the time-limited noisy Gramians (17) and (19) in Algorithm 1 is straightforward if in Step 1 there are (low-rank) factors for $\boldsymbol{P}_{\mathcal{T}} = \boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}$ and $\boldsymbol{Q}_{\mathcal{T}} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$. In LG Bayesian inference and 4D-Var, the time-limited reachability Gramian is equal to the prior covariance. Therefore, no Lyapunov equation needs to be solved. The time-limited noisy observability Gramian computation requires the solution of the modified Lyapunov equation (18) or (20). The main computational obstacle in solving the discrete or continuous modified Lyapunov equations is the computation of the matrix power $\boldsymbol{A}^{t_e}$ or the matrix exponential $e^{\boldsymbol{A} t_e}$ for the right-hand sides. There is no direct square root approach. But direct approaches quickly reach

Table 1 Time-limited (TL) Gramians in data assimilation problems

System	TL reachability Gramian	TL observability Gramian
Standard TLBT (8)	$\boldsymbol{P}_{\mathcal{T}} = \int_0^{t_e} e^{\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}e^{\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau$	$\boldsymbol{Q}_{\mathcal{T}} = \int_0^{t_e} e^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}e^{\boldsymbol{A}\tau}\mathrm{d}\tau$
	$\boldsymbol{P}_{\mathcal{T}} = \sum_{k=0}^{t_e-1}\boldsymbol{A}^k\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^k$	$\boldsymbol{Q}_{\mathcal{T}} = \sum_{k=0}^{t_e-1}(\boldsymbol{A}^{\mathrm{T}})^k\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}^k$
LG Infer. (2)	$\boldsymbol{P}_{\mathcal{T}}^{\mathrm{LG}} = \boldsymbol{\Gamma}_{\mathrm{pr}}$	$\boldsymbol{Q}_{\mathcal{T}}^{\mathrm{LG}} = \int_0^{t_e} e^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Gamma}_{\epsilon}^{-1}\boldsymbol{C}e^{\boldsymbol{A}\tau}\mathrm{d}\tau$
4D-Var (15)	$\boldsymbol{P}_{\mathcal{T}}^{\text{4D-Var}} = \boldsymbol{\Gamma}_{\mathrm{pr}}$	$\boldsymbol{Q}_{\mathcal{T}}^{\text{4D-Var}} = \sum_{k=0}^{t_e-1}(\boldsymbol{A}^{\mathrm{T}})^k\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Gamma}_{\epsilon}^{-1}\boldsymbol{C}\boldsymbol{A}^k$

their limit for high-dimensional problems. Slower iterative algorithms, like (block-)rational Krylov algorithms, may still compute a result. Shift selection is a typical bottleneck in these algorithms and they take longer than solving the Lyapunov equations for the infinite Gramians in low-dimensional settings; see, e.g., [33, 45]. Time-limited Gramians, however, allow the application of the BT Algorithm 1 to unstable systems, where Lyapunov equation solvers cannot be used.

In this section, time-limited balanced truncation has been introduced for the linear tangent model of 4D-Var and for LG Bayesian inference. Table 1 summarises the resulting Gramians for TLBT in data assimilation. By comparing the two systems in question, we have generalised the application of BT in linear Gaussian Bayesian inference to arbitrary prior covariances. We have provided a unifying definition for the observability Gramian in incremental 4D-Var. Our main contribution is that our approach can handle unstable system matrices $\boldsymbol{A}$, which are common in data assimilation problems. In Sect. 4 we will see that TLBT further outperforms the standard BT approach for short end-times.

4 Numerical Experiments for TLBT in Data Assimilation

This section is dedicated to demonstrating the performance of TLBT in linear Gaussian Bayesian inference with numerical experiments. The MATLAB code reproducing our results is available at https://github.com/joskoUP/TLBTforDA.

4.1 Comparison of Compatible and Non-Compatible Prior Covariance

In the previous section we have argued that it is reasonable to set $\boldsymbol{P}_{\mathcal{T}}^{\mathrm{LG}} = \boldsymbol{\Gamma}_{\mathrm{pr}}$ regardless of the prior-compatibility. The behaviour of non-compatible versus compatible priors in BT for LG Bayesian inference will be demonstrated by a numerical experiment. The reduced posterior statistics are given by (22). We include the optimal low-rank approach (OLR) to the posterior statistics (3) by Spantini et al. [52]:

Consider the linear Gaussian Bayesian inference problem $\boldsymbol{m} = \boldsymbol{G}\boldsymbol{x} + \boldsymbol{\epsilon}$ with forward matrix $\boldsymbol{G} \in \mathbb{R}^{nd_{\mathrm{out}}\times d}$, Fischer matrix $\boldsymbol{H} \in \mathbb{R}^{d\times d}$ and a vector of measurements $\boldsymbol{m} = [\boldsymbol{m}_1^{\mathrm{T}}, \ldots, \boldsymbol{m}_n^{\mathrm{T}}]^{\mathrm{T}}$. Prior and observation errors are assumed to be independent, centered Gaussians of the right dimensions with covariance matrices $\boldsymbol{\Gamma}_{\mathrm{pr}}$ and $\boldsymbol{\Gamma}_{\epsilon}$, respectively. The posterior statistics of $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pos}}, \boldsymbol{\Gamma}_{\mathrm{pos}})$ are approximated by $\hat{\boldsymbol{\mu}}_{\mathrm{pos}}$ and $\hat{\boldsymbol{\Gamma}}_{\mathrm{pos}}$ as follows:

Let $(\tau_i^2, \boldsymbol{v}_i)_{i=1}^d$ be the d generalised eigenvalue-eigenvector pairs of the matrix pencil $(\boldsymbol{H}, \boldsymbol{\Gamma}_{\mathrm{pr}}^{-1})$ and let $(\boldsymbol{w}_i)_{i=1}^d$ be the d generalised eigenvectors of the matrix pencil $(\boldsymbol{G}\boldsymbol{\Gamma}_{\mathrm{pr}}\boldsymbol{G}^{\mathrm{T}}, \boldsymbol{\Gamma}_{\mathrm{obs}})$. Both sequences of generalised eigenvectors are assumed to be associated with a non-increasing sequence of eigenvalues and the $\boldsymbol{w}_i$ are normalised w.r.t. $\boldsymbol{\Gamma}_{\mathrm{obs}}$.

The optimal posterior covariance approximation $\hat{\boldsymbol{\Gamma}}_{\text{pos}}$ of $\boldsymbol{\Gamma}_{\text{pos}}$ in the Förstner-metric (out of class $\mathcal{M}_r := \{\boldsymbol{\Gamma}_{\text{pr}} - \boldsymbol{K}\boldsymbol{K}^{\text{T}} \succ 0 \,|\, \text{rank}(\boldsymbol{K}) \leq r\}$) is

$$\hat{\boldsymbol{\Gamma}}_{\text{pos}} = \boldsymbol{\Gamma}_{\text{pr}} - \sum_{i=1}^{r} \frac{\tau_i^2}{1+\tau_i^2} \boldsymbol{v}_i \boldsymbol{v}_i^{\text{T}}. \tag{23a}$$

The optimal mean approximation $\hat{\boldsymbol{\mu}}_{\text{pos}}$ of $\boldsymbol{\mu}_{\text{pos}}$ in the $\|\cdot\|_{\boldsymbol{\Gamma}_{\text{pos}}^{-1}}$-norm Bayes risk (out of class $\mathcal{V}_r^{LR} := \{\boldsymbol{v} = \boldsymbol{N}\boldsymbol{m} \,|\, \boldsymbol{N} \in \mathbb{R}^{d \times n d_{\text{out}}},\ \text{rank}(\boldsymbol{N}) \leq r\}$) is

$$\hat{\boldsymbol{\mu}}_{\text{pos}} = \sum_{i=1}^{r} \frac{\tau_i}{1+\tau_i^2} \boldsymbol{v}_i \boldsymbol{w}_i^{\text{T}} \boldsymbol{m}. \tag{23b}$$

See the paper by Spantini et al. [52] for more details and explanations of the measures.

Remark 4.1 (a) For semi-positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and $(\sigma_i)_i$ the generalised eigenvalues of the matrix pencil $(\boldsymbol{A}, \boldsymbol{B})$, the Förstner-metric $d_{\mathcal{F}}(\boldsymbol{A}, \boldsymbol{B})$ is defined as $d_{\mathcal{F}}^2(\boldsymbol{A}, \boldsymbol{B}) := \text{trace}\big[\ln^2(\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{B}\boldsymbol{A}^{-\frac{1}{2}})\big] = \sum_i \ln^2(\sigma_i)$ [22].

(b) The $\|\cdot\|_{\boldsymbol{\Gamma}_{\text{pos}}^{-1}}$-norm Bayes risk $R(\delta(\boldsymbol{m}), \boldsymbol{x})$ of the measurement-based estimator $\delta(\boldsymbol{m})$ of parameter $\boldsymbol{x}$ unknown is defined as $R(\delta(\boldsymbol{m}), \boldsymbol{x}) := \mathbb{E}(\|\boldsymbol{x} - \delta(\boldsymbol{m})\|^2_{\boldsymbol{\Gamma}_{\text{pos}}^{-1}})$. The expectation is taken over the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{m}$ [39].

For the numerical experiments in this and the next subsection we have reproduced the benchmark examples by Qian et al. [46, cf. p. 22].[1] In the first presented experiment, we have modified the code for the additional computations with the non-compatible prior. We always consider LTI systems starting at $t = 0$ and running up to different end times t_e. Measurements are taken at discrete equidistant times $t_i = ih$, $(i = 1, \ldots, n)$, $t_n = t_e$. The goal is to infer the initial condition $\boldsymbol{x}_0$ by using reduced models of different ranks r. For our experiments, a true initial condition is drawn from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{pr}})$ and measurements are generated from evolving the underlying dynamical system exactly and adding $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\epsilon})$-Gaussian measurement noise. We consider a given prior non-compatible with the system matrix $\boldsymbol{A}$. Using the code from [46, Appendix D], we create a modified compatible prior.

The reduced posterior statistics (22) for this problem are computed using BT with infinite Gramians for both the given non-compatible (NC) and the modified compatible (C) prior. We compare them with the full posterior statistics (3) using the Förstner metric (for the covariance) and the $\|\cdot\|_{\boldsymbol{\Gamma}_{\text{pos}}^{-1}}$-norm Bayes risk (for the mean) derived from the true non-compatible prior (−NC) and derived with the modified compatible prior (−C). OLR for both priors and both measures is also considered. For the Bayes risk computation, we run 100 independent experiments to obtain 100 different posterior mean estimates and compute the empirical Bayes risk among them.

We study the heat equation[2] with $d = 200$, $d_{\text{out}} = 1$ and $\boldsymbol{\Gamma}_{\epsilon} = \sigma_{\text{obs}}^2$, $\sigma_{\text{obs}} = 0.008$. The non-compatible prior covariance $\boldsymbol{\Gamma}_{\text{pr}} = \boldsymbol{R}\boldsymbol{R}^{\text{T}}$ with $\boldsymbol{R} = \text{diag}(\text{ones}(1, d)) + \text{diag}(0.5 \cdot \text{randn}(1, d-1), 1) + \text{diag}(0.25 \cdot \text{randn}(1, d-2), 2)$ takes into account that neighbouring points might influence one another with decreasing intensity.

The results of the experiments are depicted in Fig. 1 Similarly to the results of Qian et al. [46], OLR (in blue/cyan) is superior to BT (in pink/orange) for the same prior. BT with the true NC prior (orange circles) performs almost as good as OLR, but the modified compatible

[1] Code taken from https://github.com/elizqian/balancing-bayesian-inference.

[2] LTI system matrices and documentation at http://slicot.org/20-site/126-benchmark-examples-for-model-reduction.

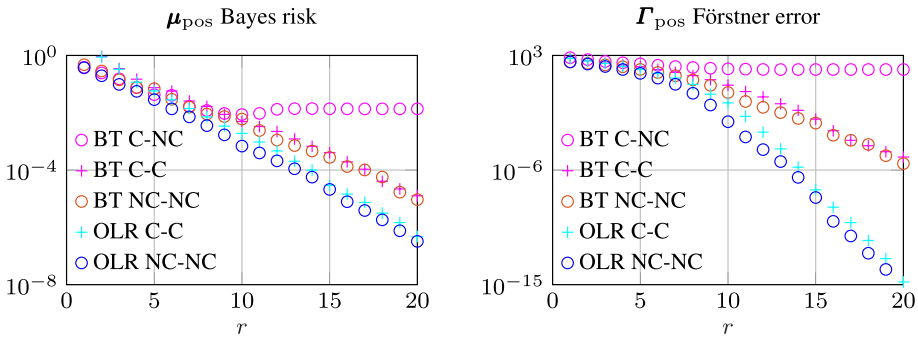

Fig. 1 Comparison of a given non-compatible prior (NC) and a modified compatible prior (C) in balanced truncation (BT) [46] and the optimal low-rank approach (OLR) [52] of LG Bayesian inference for the heat equation. The first letters denote the prior used for the approx. posterior computation and the second the one used for the reference. Measurements are spaced $h = 0.005$ apart inside $\mathcal{T} = [0, 10]$

prior gives bad approximations in the original (NC) measures (pink circles). Considering these approximations in the measures with regard to the modified compatible prior (C), they are almost as good as the original ones (pink and cyan crosses). It is thus not useful to enforce prior-compatibility for a non-compatible prior containing valuable knowledge as the modification changes the results significantly. Compatible priors, however, perform well in the measures derived from them. Building compatible priors may therefore be a way of constructing a prior that fits the system dynamics when good prior knowledge is not available otherwise.

Non-compatible priors are particularly important when dealing with unstable system matrices $\boldsymbol{A}$. Time-limited BT is especially suitable for such systems in data assimilation, as will be demonstrated in the next experiments.

4.2 Comparison of TLBT and Standard BT with Infinite Gramians

We apply TLBT to benchmark examples and compare the approximate posterior statistics with full balancing (BT) and with the optimal low-rank approach (OLR). For the experiments in this section we consider the same setup as before with added code for the time-limited Gramian computation and with the following change: $\boldsymbol{\Gamma}_{\text{pr}}$ is obtained as the solution of $\boldsymbol{A}\boldsymbol{\Gamma}_{\text{pr}} + \boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{A}^{\text{T}} = -\boldsymbol{B}\boldsymbol{B}^{\text{T}}$ for $\boldsymbol{B}$ according to the experiment; see [46]. The TLBT-reduced posterior statistics (22) are computed as described in Sect. 3.2

The true posterior statistics (3) are compared in the $\|\cdot\|_{\boldsymbol{\Gamma}_{\text{pos}}^{-1}}$-norm Bayes risk and the Förstner-metric with $(\boldsymbol{\mu}_{\text{pos,TLBT}}, \boldsymbol{\Gamma}_{\text{pos,TLBT}})$, $(\boldsymbol{\mu}_{\text{pos,BT}}, \boldsymbol{\Gamma}_{\text{pos,BT}})$, and $(\hat{\boldsymbol{\mu}}_{\text{pos}}, \hat{\boldsymbol{\Gamma}}_{\text{pos}})$ from OLR. The time-limited Gramians for TLBT are computed by the rational Krylov method by Kürschner [33, p. 1830].[3] We have to choose parameters: the approximation tolerance (tol) for the low-rank solution of $f(\boldsymbol{A})\boldsymbol{B}$ and the maximum number of rational Krylov iterations (maxit). The Lyapunov equation for the recent low-rank approximation is solved at every s-th step. We studied both the heat equation and the ISS1R module.[2] All parameters are given in Table 2.

The results depicted in Fig. 2 for the heat equation and in Fig. 3 for the ISS model are as expected: for low end times, especially $t_e = 1$, TLBT performs significantly better than BT with infinite Gramians. This is caused by the faster decay of the Hankel singular values [33].

[3] Code from https://zenodo.org/record/7366026.

Table 2 Parameters for the LTI examples

Example	d	d_{out}	$\boldsymbol{B}$	$\boldsymbol{\Gamma}_\epsilon$	s	Tol	Maxit
Heat equation	200	1	$\boldsymbol{I}_d$	$[0.008^2]$	5	10^{-8}	100
ISS1R module	270	3	See[4]	$\begin{bmatrix} 0.0025^2 & 0 & 0 \\ 0 & 0.0005^2 & 0 \\ 0 & 0 & 0.0005^2 \end{bmatrix}$	3	10^{-4}	100

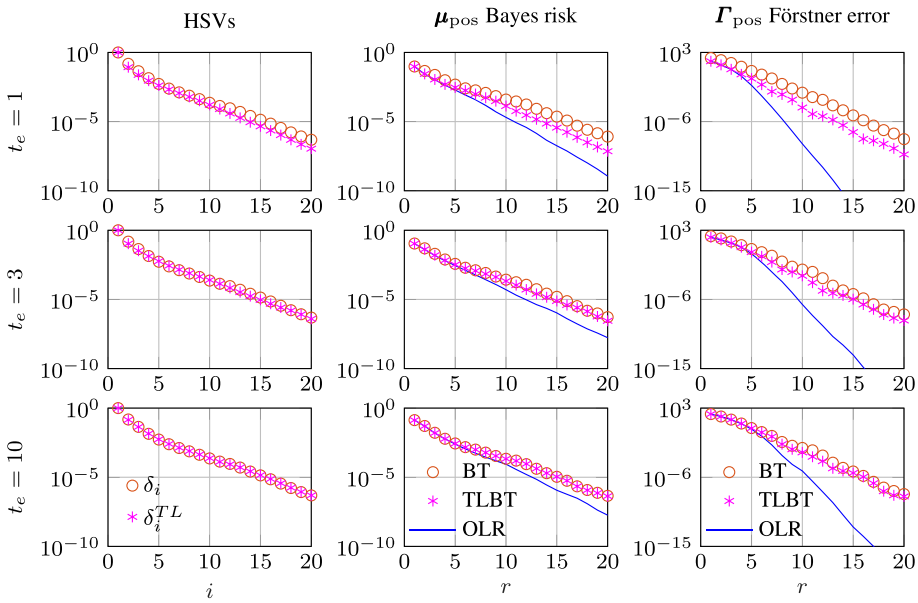

Fig. 2 Comparison of time-limited balanced truncation (TLBT) with standard balanced truncation (BT) [46] and the optimal low-rank approach (OLR) [52] of LG Bayesian inference for the heat equation model. Measurements are spaced $h = 0.005$ apart inside $\mathcal{T} = [0, t_e]$ for three different end times $t_e = 1, 3, 10$. In the left panel, the normalised square roots δ_i and δ_i^{TL} of the generalised eigenvalues of the matrix pencils $(\boldsymbol{Q}_\infty^{\text{LG}}, \boldsymbol{\Gamma}_{\text{pr}}^{-1})$ and $(\boldsymbol{Q}_{\mathcal{T}}^{\text{LG}}, \boldsymbol{\Gamma}_{\text{pr}}^{-1})$ are plotted, corresponding to the Hankel singular values

With increasing end time t_e the TLBT approximation results approach the BT approximation results. For the heat equation, OLR outperforms TLBT, whereas for the ISS model, optimality can be attained by TLBT, especially for lower reduction ranks r. For low end times, TLBT is able to reach the optimal OLR approximation even when standard BT cannot. This is a huge advantage of using time-limited Gramians. Especially in the ISS1R model for $t_e = 1$ this behaviour can be observed.

OLR is superior to TLBT regarding approximation quality, but not regarding speed. The main advantage of (TL)BT over OLR is that $\boldsymbol{G}$ and $\boldsymbol{H}$ are computed in reduced versions of lower dimensionality. They are obtained by forwarding a reduced rather than a full dynamical system (2). Computations based on a reduced forward operator $\boldsymbol{G}$ run faster. This is particulary relevant during the online computation of multiple posterior means for a series of measurements. OLR is optimal for linear Gaussian Bayesian inference. For BT, there are generalisations to time-varying [50] and nonlinear [4, 31, 32] systems that should be further explored in the context of data assimilation.

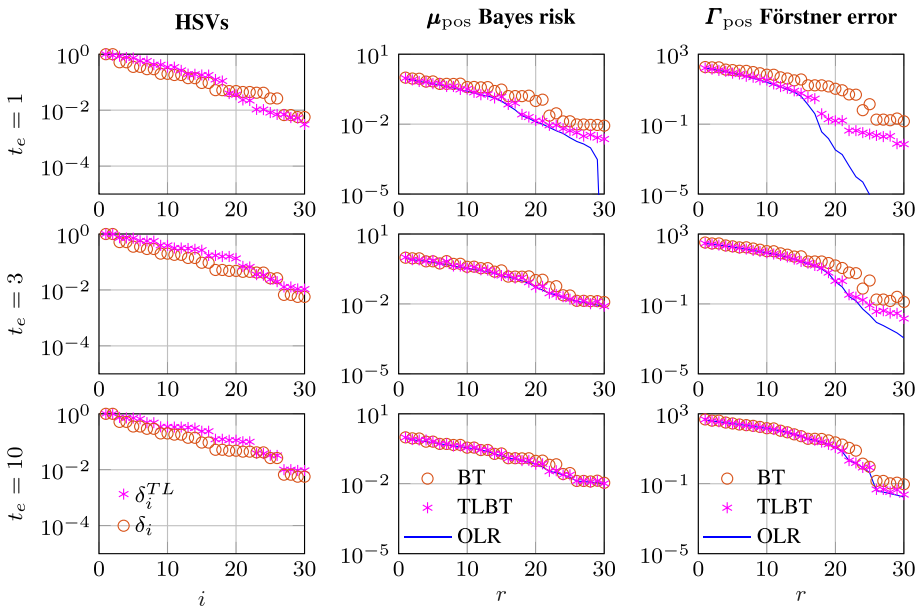

Fig. 3 Quantities plotted as in Fig. 2; here for the ISS1R model. Measurements are equispaced with $h = 0.1$ inside $\mathcal{T} = [0, t_e]$ for three different end times $t_e = 1, 3, 10$

4.3 TLBT for Unstable Systems

Time-limited BT for unstable systems is applied to advection and diffusion for an instantaneous point release. For simplicity, we assume the system is one-dimensional and consider isotropic and homogeneous diffusion. For $K(z, t)$, the location- and time-dependent concentration, the process is described by

$$\frac{\partial K}{\partial t} = D\frac{\partial^2 K}{\partial z^2} - u\frac{\partial K}{\partial z}.$$

The diffusion coefficient is $D = 0.02$ and the mean flow is $u = 0.01$. The system is discretised using finite differences for state dimensions $d = 200$ and $d = 1200$, respectively. The system matrix $\boldsymbol{A}$ has both positive and negative eigenvalues and is unstable. The observation operator $\boldsymbol{C}$ with $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} = \left[\frac{1}{d} \dots \frac{1}{d}\right]\boldsymbol{x}$ gives the mean concentration and $d_{\text{out}} = 1$. The observation error covariance is set to $\boldsymbol{\Gamma}_\epsilon = \sigma_{\text{obs}}^2$, $\sigma_{\text{obs}} = 0.008$. The prior covariance is chosen to be $\boldsymbol{\Gamma}_{\text{pr}} = \boldsymbol{I}_d$ and is used as time-limited reachability Gramian, $\boldsymbol{P}_{\mathcal{T}}^{\text{LG}} = \boldsymbol{\Gamma}_{\text{pr}}$. This Gramian is not prior-compatible, i.e., $\boldsymbol{A}\boldsymbol{\Gamma}_{\text{pr}} + \boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{A}^{\text{T}} = \boldsymbol{A} + \boldsymbol{A}^{\text{T}}$ is not negative-semidefinite, but as discussed in Sect. 3.2.1 it is not necessary to assume prior-compatibility.

After computing the observability Gramian $\boldsymbol{Q}_{\mathcal{T}}^{\text{LG}}$ (see Table 1), the algorithm is the same as for stable systems. To account for the instability of the system, which makes it unsuitable for long-term predictions, we use smaller end times than in the stable examples. The BT approach with infinite Gramians is not applicable to unstable systems. We, therefore, use a different approach by Qian et al. [46] for comparison: the Fisher matrix $\boldsymbol{H}$ is set as the observability Gramian. No Lyapunov equation solution is required, but the computation of the full Fisher matrix is expensive. We refer to this approach as BT-H.

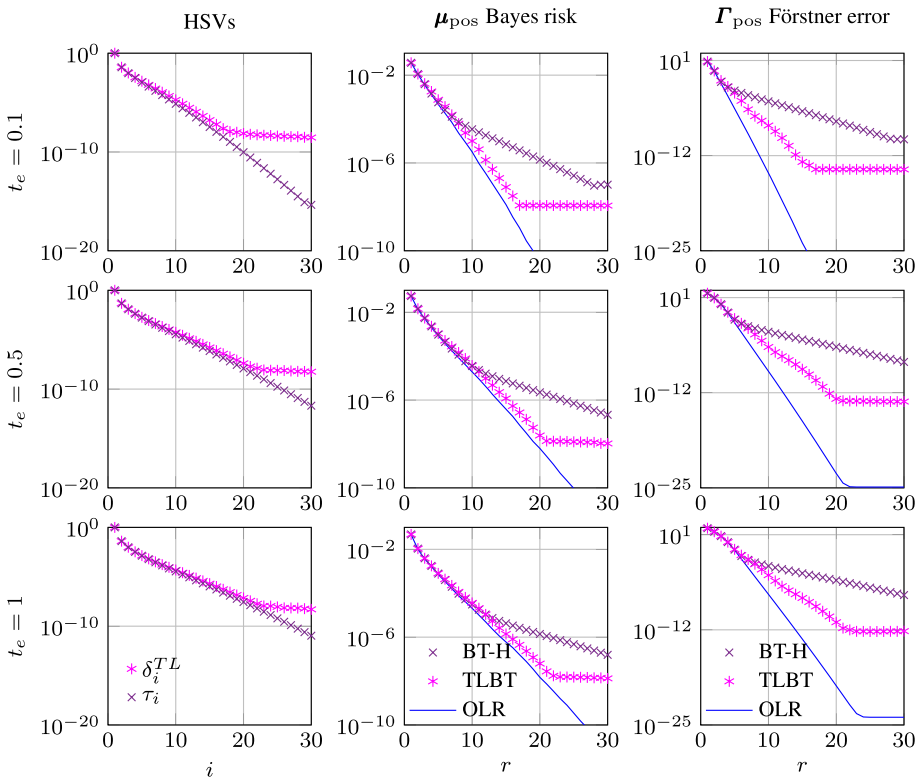

Fig. 4 Comparison of time-limited balanced truncation (TLBT) and balanced truncation with Fisher matrix $\boldsymbol{H}$ as observability Gramian (BT-H) [46] and the optimal low-rank approach (OLR) [52] of LG Bayesian inference for the unstable discretised advection–diffusion equation; $d = 200$. Measurements are spaced $h = 0.001$ apart inside $\mathcal{T} = [0, t_e]$ for three different end times $t_e = 0.1, 0.5, 1$

Figure 4 illustrates that in our case with measurements spread quite far apart ($h = 0.001$), TLBT surpasses BT-H. With higher measurement frequency, the Fisher matrix approaches the time-limited observability Gramian and the results become similar. OLR ignores system stability and can be applied regardless of the system matrix $\boldsymbol{A}$. As demonstrated in Fig. 4, TLBT recovers the optimal posterior mean Bayesian risk well, but OLR outperforms TLBT for the posterior covariance approximation in the Förstner metric. The error rates of TLBT (better than 10^{-12}) are still impressive and sufficient for applications.

For Fig. 5 we increase the system dimension to $d = 1200$. This introduces positive eigenvalues λ_+ into the unstable system matrix $\boldsymbol{A}$, which are about an order of magnitude higher than before. The increasing instability affects the numerical procedure: all posterior predictions become less reliable, especially for the largest end time $t_e = 1$. This is caused by the blow-up of the values of $e^{\lambda_+ \tau}$ in $e^{\boldsymbol{A}\tau}$ for $\boldsymbol{Q}_{\mathcal{T}}^{\text{LG}}$, already for moderate end times t_e. In the mean prediction, TLBT outperforms the state-of-the-art BT-H, which is not robust to instabilities since it uses exponentials of $\boldsymbol{A}$. OLR mean prediction also fails for unstable systems and only achieves the same level of error as TLBT. These results emphasise the importance of using a time-limited approach and short observation intervals for unstable systems.

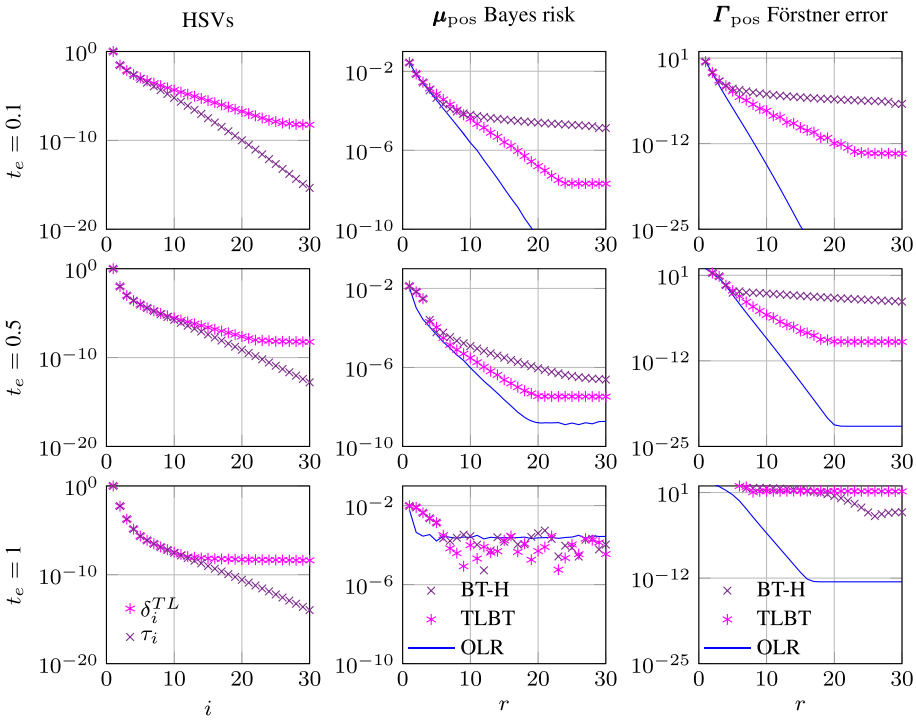

Fig. 5 Model and plotted quantities as in Fig. 4; here for $d = 1200$. Measurements are equispaced with $h = 0.001$ inside $\mathcal{T} = [0, t_e]$ for three different end times $t_e = 0.1, 0.5, 1$

The generalisation of BT to TLBT works well, but has a main disadvantage: the numerical algorithms require the solution of a Lyapunov equation. This is only numerically feasible for (anti)stable $\boldsymbol{A}$. In TLBT for unstable systems, the integral for $\boldsymbol{Q}_{\mathcal{T}}^{\text{LG}}$ has to be approximated or calculated according to its definition [33]. To make TLBT faster and more generally applicable to large problems, further work in this direction is needed.

In this section we have demonstrated how TLBT extends the idea of Gramian-based model reduction for Bayesian inference to unstable systems. For short observation intervals TLBT significantly outperforms BT with infinite Gramians, despite the increased numerical cost of approximating the posterior statistics.

5 Conclusion

Model order reduction is central to data assimilation because models and data are typically high-dimensional and expensive to process. To this end, the adaptation of model reduction techniques from systems theory to data assimilation problems is essential. This work has further strengthened the link between the two research areas.

Time-limited balanced truncation (TLBT) enabled us to significantly generalise the theory of balancing Bayesian inference to systems with an unstable system matrix $\boldsymbol{A}$ and non-compatible priors. TLBT is, thus, particularly suited to data assimilation applications such as numerical weather prediction, where short-term forecasts are required for unstable systems

[12, 13]. One of the popular techniques is incremental 4D-Var and TLBT is perfectly tailored to model reduction of unstable linear tangent models in this context. The numerical efficiency of TLBT needs further investigation, e.g., by using more advanced rational Krylov methods.

Interesting research questions in data assimilation include inhomogeneous initial conditions, errors in the dynamical forward model or time-varying systems. Similarities with systems theory may be instructive. Literature on such problems in the deterministic setting is available, but not extensive; see, e.g., [3, 29, 34, 50].

Although our approach of using TLBT in data assimilation is more general than the existing ones, it is so far limited to linear Gaussian Bayesian inference. The next step in the exploration of model order reduction for Bayesian inference is the generalisation to nonlinear settings. The 4D-Var method could be a step in the right direction, as it already deals with nonlinear operators. Concepts from systems theory for the reduction of nonlinear dynamical systems may also prove useful here. This opens up a broad field of investigation at the interface between the two research communities.

Acknowledgements We thank Patrick Kürschner for providing his code for time-limited balanced truncation and helpful insights on this topic. We also thank Elizabeth Qian for the fruitful discussions about her work and during her visit to the University of Potsdam. We thank the reviewers and Thomas Mach for critically reading the manuscript and suggesting substantial improvements. Our research has been partially funded by the Deutsche Forschungsgemeinschaft (DFG)—Project-ID 318763901–SFB1294.

Author Contributions All authors contributed to the research conception and design. Conceptual reflections and numerical experiments were performed by JK. The first draft of the manuscript was written by JK and all authors commented on previous versions of the manuscript. After reading, all authors approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Our research has been partially funded by the Deutsche Forschungsgemeinschaft (DFG)—Project-ID 318763901–SFB1294.

Data availability The MATLAB code reproducing our results is available at the *TLBTforDA* repository, https://github.com/joskoUP/TLBTforDA.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Antoulas, A.C.: Approximation of Large-Scale Dynamical Systems. SIAM, Philadelphia (2005). https://doi.org/10.1137/1.9780898718713
2. Antoulas, A.C., Beattie, C.A., Gugercin, S.: Interpolatory Methods for Model Reduction. SIAM, Philadelphia (2020). https://doi.org/10.1137/1.9781611976083
3. Beattie, C., Gugercin, S., Mehrmann, V.: Model reduction for systems with inhomogeneous initial conditions. Syst. Control Lett. **99**, 99–106 (2017). https://doi.org/10.1016/j.sysconle.2016.11.007

4. Benner, P., Goyal, P.: Balanced truncation model order reduction for quadratic-bilinear control systems. arXiv:1705.00160 (2017). https://doi.org/10.48550/ARXIV.1705.00160
5. Benner, P., Saak, J.: Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey. GAMM-Mitt. **36**(1), 32–52 (2013). https://doi.org/10.1002/gamm.201310003
6. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Rev. **57**(4), 483–531 (2015). https://doi.org/10.1137/130932715
7. Benner, P., Ohlberger, M., Cohen, A., Willcox, K.: Model Reduction and Approximation. SIAM, Philadelphia (2017). https://doi.org/10.1137/1.9781611974829
8. Benner, P., Qiu, Y., Stoll, M.: Low-rank eigenvector compression of posterior covariance matrices for linear Gaussian inverse problems. SIAM-ASA J. Uncertain. Quantif. **6**(2), 965–989 (2018). https://doi.org/10.1137/17M1121342
9. Benner, P., Goyal, P., Kramer, B., Peherstorfer, B., Willcox, K.: Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms. Comput. Methods Appl. Mech. Eng. **372**, 113433 (2020). https://doi.org/10.1016/j.cma.2020.113433
10. Bernstein, D.S., Davis, L.D., Hyland, D.C.: The optimal projection equations for reduced-order, discrete-time modeling, estimation, and control. J. Guid. Control Dyn. **9**(3), 288–293 (1986). https://doi.org/10.1016/0167-6911(87)90106-X
11. Boess, C.: Using model reduction techniques within the incremental 4D-Var method. Ph.D. thesis, Universität Bremen (2008). https://media.suub.uni-bremen.de/bitstream/elib/2608/1/00011290.pdf
12. Boess, C., Lawless, A., Nichols, N., Bunse-Gerstner, A.: State estimation using model order reduction for unstable systems. Comput. Fluids **46**(1), 155–160 (2011). https://doi.org/10.1016/j.compfluid.2010.11.033
13. Bonavita, M., Lean, P.: 4D-Var for numerical weather prediction. Weather **76**(2), 65–66 (2021). https://doi.org/10.1002/wea.3862
14. Carlin, B.P., Louis, T.A.: Bayesian Methods for Data Analysis. Chapman and Hall/CRC, New York (2009). https://doi.org/10.1201/b14884
15. Carrassi, A., Bocquet, M., Bertino, L., Evensen, G.: Data assimilation in the geosciences: an overview of methods, issues, and perspectives. Wiley Interdiscip. Rev. Clim. Change **9**(5), e535 (2018). https://doi.org/10.1002/wcc.535
16. Courtier, P., Thépaut, J.N., Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteorol. Soc. **120**(519), 1367–1387 (1994). https://doi.org/10.1002/qj.49712051912
17. Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification, pp. 311–428. Springer, Cham (2017). https://link.springer.com/referenceworkentry/10.1007/978-3-319-12385-1_7
18. Dee, D.P.: Simplification of the Kalman filter for meteorological data assimilation. Q. J. R. Meteorol. Soc. **117**(498), 365–384 (1991). https://doi.org/10.1002/qj.49711749806
19. Duff, I.P., Goyal, P., Benner, P.: Balanced truncation for a special class of bilinear descriptor systems. IEEE Control Syst. Lett. **3**(3), 535–540 (2019). https://doi.org/10.1109/LCSYS.2019.2911904
20. Engbert, R., Rabe, M.M., Kliegl, R., Reich, S.: Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics. Bull. Math. Biol. **83**, 1–19 (2020). https://doi.org/10.1007/s11538-020-00834-8
21. Fletcher, S.J.: Data Assimilation Geoscience. Elsevier, Amsterdam (2017). https://doi.org/10.1016/B978-0-12-804444-5.09996-7
22. Förstner, W., Moonen, B.: A metric for covariance matrices. In: Grafarend, E.W., Krumm, F.W., Schwarze, V.S. (eds.) Geodesy-The Challenge of the 3rd Millennium, pp. 299–309. Springer, Berlin, Heidelberg (2003)
23. Freitag, M.A.: Numerical linear algebra in data assimilation. GAMM-Mitt. **43**(3), e202000014 (2020)
24. Gawronski, W., Juang, J.N.: Model reduction in limited time and frequency intervals. Internat. J. Syst. Sci. **21**(2), 349–376 (1990). https://doi.org/10.1080/00207729008910366
25. Ghil, M.: Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. Dyn. Atmos. Oceans **13**(3), 171–218 (1989). https://doi.org/10.1016/0377-0265(89)90040-7
26. Ghil, M., Malanotte-Rizzoli, P.: Data assimilation in meteorology and oceanography. In: Dmowska, R., Saltzman, B. (eds.) Advances in Geophysics, pp. 141–266. Elsevier, Amsterdam (1991)
27. Green, D.: Model order reduction for large-scale data assimilation problems. Ph.D. thesis, University of Bath (2019). https://researchportal.bath.ac.uk/en/studentTheses/model-order-reduction-for-large-scale-data-assimilation-problems
28. Gugercin, S., Antoulas, A.C.: A survey of model reduction by balanced truncation and some new results. Internat. J. Control **77**(8), 748–766 (2004). https://doi.org/10.1080/00207170410001713448

29. Heinkenschloss, M., Reis, T., Antoulas, A.: Balanced truncation model reduction for systems with inhomogeneous initial conditions. Automatica **47**(3), 559–564 (2011). https://doi.org/10.1016/j.automatica.2010.12.002
30. Kostelich, E.J., Kuang, Y., McDaniel, J.M., Moore, N.Z., Martirosyan, N.L., Preul, M.C.: Accurate state estimation from uncertain data and models: an application of data assimilation to mathematical models of human brain tumors. Biol. Direct **6**, 64 (2011). https://doi.org/10.1186/1745-6150-6-64
31. Kramer, B., Gugercin, S., Borggaard, J., Balicki, L.: Nonlinear balanced truncation: Part 1-Computing energy functions. arXiv:2209.07645 (2022). https://doi.org/10.48550/arXiv.2209.07645
32. Kramer, B., Gugercin, S., Borggaard, J.: Nonlinear balanced truncation: Part 2 – Model reduction on manifolds. arXiv:2302.02036 (2023). https://doi.org/10.48550/arXiv.2302.02036
33. Kürschner, P.: Balanced truncation model order reduction in limited time intervals for large systems. Adv. Comput. Math. **44**(6), 1821–1844 (2018). https://doi.org/10.1007/s10444-018-9608-6
34. Lang, N., Saak, J., Stykel, T.: Balanced truncation model reduction for linear time-varying systems. Math. Comput. Model. Dyn. Syst. **22**(4), 267–281 (2016). https://doi.org/10.1080/13873954.2016.1198386
35. Law, K., Stuart, A., Zygalakis, K.: Data Assimilation: a Mathematical Introduction. Texts in Applied Mathematics, Springer, Cham (2015)
36. Lawless, A.S., Gratton, S., Nichols, N.K.: An investigation of incremental 4D-Var using non-tangent linear models. Q. J. R. Meteorol. Soc. **131**(606), 459–476 (2005). https://doi.org/10.1256/qj.04.20
37. Lawless, A.S., Nichols, N.K., Boess, C., Bunse-Gerstner, A.: Approximate Gauss-Newton methods for optimal state estimation using reduced-order models. Int. J. Numer. Methods Fluids **56**(8), 1367–1373 (2008). https://doi.org/10.1002/fld.1629
38. Lawless, A.S., Nichols, N.K., Boess, C., Bunse-Gerstner, A.: Using model reduction methods within incremental four-dimensional variational data assimilation. Mon. Weather Rev. **136**(4), 1511–1522 (2008). https://doi.org/10.1175/2007MWR2103.1
39. Lehmann, E.L., Casella, G.: Theory of Point Estimation. Springer, New York (1998). https://doi.org/10.1007/b98854
40. Mahalanobis, P.C.: On the generalized distance in statistics. Proc. Natl. Acad. Sci. **2**, 49–55 (1936)
41. Moore, B.: Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Trans. Automat. Control **26**(1), 17–32 (1981). https://doi.org/10.1109/TAC.1981.1102568
42. Mullis, C., Roberts, R.: Synthesis of minimum roundoff noise fixed point digital filters. IEEE Trans. Circuits Syst. **23**(9), 551–562 (1976). https://doi.org/10.1109/TCS.1976.1084254
43. Park, S.K., Xu, L.: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-43415-5
44. Peherstorfer, B., Willcox, K.: Data-driven operator inference for nonintrusive projection-based model reduction. Comput. Methods Appl. Mech. Eng. **306**, 196–215 (2016). https://doi.org/10.1016/j.cma.2016.03.025
45. Pontes Duff, I., Kürschner, P.: Numerical computation and new output bounds for time-limited balanced truncation of discrete-time systems. Linear Algebra Appl. **623**, 367–397 (2021). https://doi.org/10.1016/j.laa.2020.09.029
46. Qian, E., Tabeart, J.M., Beattie, C., Gugercin, S., Jiang, J., Kramer, P.R., Narayan, A.: Model reduction for linear dynamical systems via balancing for Bayesian inference. J. Sci. Comput. (2022). https://doi.org/10.1007/s10915-022-01798-8
47. Redmann, M.: An L_T^2-error bound for time-limited balanced truncation. Syst. Control Lett. **136**, 104620 (2020). https://doi.org/10.1016/j.sysconle.2019.104620
48. Redmann, M., Freitag, M.A.: Balanced model order reduction for linear random dynamical systems driven by Lévy noise. J. Comput. Dyn. **5**(1 & 2), 33–59 (2018). https://doi.org/10.3934/jcd.2018002
49. Redmann, M., Kürschner, P.: An output error bound for time-limited balanced truncation. Syst. Control Lett. **121**, 1–6 (2018). https://doi.org/10.1016/j.sysconle.2018.08.004
50. Sandberg, H., Rantzer, A.: Balanced truncation of linear time-varying systems. IEEE Trans. Automat. Control **49**(2), 217–229 (2004). https://doi.org/10.1109/TAC.2003.822862
51. Sasaki, Y.: Some basic formalisms in numerical variational analysis. Mon. Weather Rev. **98**, 875–883 (1970)
52. Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L., Marzouk, Y.: Optimal low-rank approximations of Bayesian linear inverse problems. SIAM J. Sci. Comput. **37**(6), A2451–A2487 (2015). https://doi.org/10.1137/140977308
53. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia (2005). https://doi.org/10.1137/1.9780898717921

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.