ON OPTIMAL POINTWISE IN TIME ERROR BOUNDS AND DIFFERENCE QUOTIENTS FOR THE PROPER ORTHOGONAL DECOMPOSITION*

BIRGUL KOC†, SAMUELE RUBINO‡, MICHAEL SCHNEIER§, JOHN SINGLER¶, AND TRAIAN ILIESCU†

Abstract. In this paper, we resolve several long-standing issues dealing with optimal pointwise in time error bounds for proper orthogonal decomposition (POD) reduced order modeling of the heat equation. In particular, we study the role played by difference quotients (DQs) in obtaining reduced order model (ROM) error bounds that are optimal with respect to both the time discretization error and the ROM discretization error. When the DQs are not used, we prove that both the POD projection error and the ROM error are suboptimal. When the DQs are used, we prove that both the POD projection error and the ROM error are optimal. The numerical results for the heat equation support the theoretical results.

1. Introduction. In this paper, we consider the one-dimensional heat equation

$$(1.1) \qquad u_t - \nu\, u_{xx} = f\,,$$

where the spatial domain is $[0, 1]$, the time domain is $[0, T]$, and ν is the diffusion coefficient. For simplicity, we consider homogeneous Dirichlet boundary conditions $u(0, t) = u(1, t) = 0$ for $t > 0$ and given initial conditions $u(x, 0) = u_0(x)$.

We also consider projection reduced order models (ROMs) for the heat equation. Specifically, we consider the proper orthogonal decomposition (POD) [23], which can be summarized as follows: (i) The full order model (FOM) for (1.1) is run for selected parameter values and/or time intervals to generate a set of snapshots $\{u^0, u^1, \ldots, u^N\}$; (ii) these snapshots and the singular value decomposition are used to construct an

†Department of Mathematics, Virginia Tech, Blacksburg, VA 24061 USA (birgul@vt.edu, iliescu@vt.edu).

‡Departamento EDAN & IMUS, Universidad de Sevilla, 41004 Seville, Spain (samuele@us.es).

§Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260 USA (mhs64@pitt.edu).

¶Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409 USA (singlerj@mst.edu).

orthonormal ROM basis $\{\varphi_1, \ldots, \varphi_s\}$ for a Hilbert space $\mathcal{H}$, where s is the rank of the snapshot matrix; (iii) the ROM approximation

$$(1.2) \qquad u(x, t_n) \approx u_r^n(x) = \sum_{j=1}^{r} u_j^n \, \varphi_j(x), \qquad n = 1, \ldots, N,$$

where $r < s$ is the ROM dimension, is used together with a Galerkin projection and a time discretization to yield a system of equations for u_j^n, which are the sought ROM coefficients.

DEFINITION 1.1 (generic constant C). *For clarity, in what follows, we will denote by C a generic positive constant that may vary from one line to another but which is always independent of the discretization parameters.*

In the pioneering paper [34], Kunisch and Volkwein laid the foundations of numerical analysis for POD (see, e.g., [35, 40, 44] for relevant work). In particular, for the ROM error

$$(1.3) \qquad e^n(x) = u(x, t_n) - u_r^n(x), \qquad n = 1, \ldots, N,$$

they proved the following error bound (see Theorem 7 in [34]):

$$(1.4) \quad \frac{1}{N+1} \sum_{n=1}^{N} \|e^n\|_{L^2}^2 \le C \left(\text{time discretization error} + \text{ROM discretization error} \right).$$

This estimate was later extended to include the spatial discretization error and a pointwise in time estimate in [29] (see, e.g., [33, 44] for alternative pointwise in time estimates), i.e.,

$$\|e^n\|_{L^2} \le C \left(\text{space discretization error} + \text{time discretization error} \right.$$

$$(1.5) \qquad \left. + \text{ROM discretization error} \right).$$

Estimate (1.5) relied on an assumption about the POD projection error, which roughly says that the POD projection error at each time step is of the same order as the POD projection error at the remaining time steps. This assumption has since been generally used in proving pointwise in time error bounds for parabolic equations.

We emphasize that the error bound (1.5) includes all three ROM error sources: (i) the space discretization error, which results from the spatial discretization of the heat equation (1.1) with classical numerical methods, e.g., finite elements (FEs); (ii) the time discretization error, which results from the time discretization of the heat equation (1.1) with classical numerical methods, e.g., Euler or Crank–Nicolson (CN) methods; and (iii) the ROM discretization error, which results from the truncation in (1.2).

Pointwise error bounds like (1.5) without the ROM discretization error are standard in the numerical analysis of classical discretization methods for the heat equation (see, e.g., [51]). Pointwise errors bounds such as (1.5) are desirable since they eliminate the possibility of an error "spike" at a certain point in time, which in principle is allowed with a bound of the form (1.4). Our goal here is to better understand these pointwise error bounds in the context of POD reduced order modeling.

A fundamental issue in the POD numerical analysis is the *optimality* of the error bound (1.5). We emphasize that there are three types of optimality, corresponding to the three types of discretization levels: (i) space discretization optimality, (ii) time discretization optimality, and (iii) ROM discretization optimality. We discuss each optimality type below.

Space discretization optimality. For simplicity, we consider an FE spatial discretization. We emphasize, however, that other standard numerical methods (e.g., finite difference, spectral, or spectral element methods) could be considered. An error bound is optimal with respect to the spatial discretization if the error scalings with respect to the spatial discretization parameters only are of the following form

$$\|e^n\|_{L^2} = \mathcal{O}(h^{m+1}),$$
(1.6)

$$\|\nabla e^n\|_{L^2} = \mathcal{O}(h^m),$$
(1.7)

where h is the size of the FE mesh and m is the FE order. Proving estimates that are optimal with respect to the spatial discretization is relatively straightforward (see, e.g., [14, 27, 29]), since it follows the standard FE numerical analysis [51]. Thus, the spatial discretization error component is generally ignored in POD numerical analysis papers (see, e.g., [34, 54]). To simplify the presentation, we will not discuss the spatial discretization optimality in this paper. We note, however, that our results can be extended in a straightforward manner to include the spatial discretization optimality.

Time discretization optimality. An error bound is optimal with respect to the time discretization if the error scalings with respect to the time discretization parameters only are of the following form:

$$\|e^n\|_{L^2} = \mathcal{O}(\Delta t^k),$$
(1.8)

where Δt is the time step size used in the time discretization, and k is the time discretization order (e.g., $k = 1$ for Euler's method, and $k = 2$ for CN).

The importance of the time discretization optimality was recognized early on. In Remark 1 of [34], Kunisch and Volkwein proposed the *difference quotients (DQs)* (i.e., scaled snapshots of the form $(u^n - u^{n-1})/\Delta t$, $n = 1, \ldots, N$) as a means to achieve time discretization optimality. Specifically, on page 121 of [34], the authors noted that, in the DQ case (i.e., if the DQs are used to build the POD basis), time discretization optimal error bounds of the type (1.8) follow. However, in the noDQ case (i.e., if the DQs are not used), the norm squared error bound has a suboptimal (Δt^{-2}) factor.

A major development in the study of POD optimality was made by Chapelle, Gariah, and Sainte-Marie in [5]. The authors showed that using the L^2 projection instead of the Ritz projection used in [34] (which is standard in FE numerical analysis [51, 55]) avoids the difficulties posed by the POD approximation of the time derivative. Pointwise error bounds were not considered in [5]; however, it can be checked that using the L^2 projection eliminates the need to use DQs to achieve time discretization optimality if the pointwise POD projection error assumption mentioned earlier is made.

ROM discretization optimality. The first discussion of the ROM discretization optimality was presented in [28]. In that work, a pointwise in time error bound was said to be optimal with respect to the ROM discretization if the error scalings with respect to the ROM discretization parameters only take one of the following forms:

$$(1.9) \quad \|e^n\|_{L^2}^2 = \mathcal{O}\left(\frac{1}{N+1} \sum_{n=0}^{N} \|\eta^{proj}(t_n)\|_{L^2}^2 \right) = \mathcal{O}\left(\sum_{i=r+1}^{s} \lambda_i \right),$$

$$(1.10) \quad \|\nabla e^n\|_{L^2}^2 = \mathcal{O}\left(\frac{1}{N+1} \sum_{n=0}^{N} \|\nabla\eta^{proj}(t_n)\|_{L^2}^2 \right) = \mathcal{O}\left(\sum_{i=r+1}^{s} \lambda_i \|\nabla\varphi_i\|_{L^2}^2 \right),$$

where η^{proj} is the *POD projection error*, which is defined as

$$(1.11) \qquad \eta^{proj}(x,t) = u(x,t) - \sum_{i=1}^{r} \Big(u(\cdot,t), \varphi_i(\cdot) \Big)_{\mathcal{H}} \varphi_i(x),$$

and λ_i and φ_i are POD eigenvalues and modes. The first significant development in the study of POD optimality was made in [28], where it was shown (utilizing the technique from [5]) that not using the DQs yields pointwise error bounds that are suboptimal with respect to the ROM discretization. (We note that the optimality with respect to the time discretization was not considered in [28].) Specifically, in the noDQ case, it was shown in [28] that

$$(1.12) \quad \|e^n\|_{L^2}^2 = \mathcal{O}\left(\frac{1}{N+1} \sum_{n=0}^{N} \|\nabla\eta^{proj}(t_n)\|_{L^2}^2 \right) = \mathcal{O}\left(\sum_{i=r+1}^{s} \lambda_i \|\nabla\varphi_i\|_{L^2}^2 \right),$$

which is suboptimal with respect to the ROM discretization. Furthermore, in the DQ case, it was shown in [28] that

$$(1.13) \qquad \|e^n\|_{L^2}^2 = \mathcal{O}\left(\frac{1}{N+1} \sum_{n=0}^{N} \|\eta^{proj}(t_n)\|_{L^2}^2 \right) = \mathcal{O}\left(\sum_{i=r+1}^{s} \lambda_i \right),$$

which is optimal with respect to the ROM discretization. However, two assumptions on the POD projection errors were made in order to establish these results.

To summarize, the current state of the art in POD optimality *suggests* that

$$(1.14) \qquad \boxed{\text{DQs are needed for optimal POD error bounds.}}$$

We emphasize that, to our knowledge, (1.14) *has never been proved.* Indeed, [34] focused on the time discretization optimality but ignored the ROM discretization optimality. Specifically, the authors proved that using DQs yields error bounds that are optimal with respect to the time discretization, but not necessarily with respect to the ROM discretization. In [5], the authors considered the noDQ case and developed a framework that yields error bounds that are optimal with respect to the time discretization, but not necessarily with respect to the ROM discretization. A completely different approach was taken in [28], where the focus was on ROM discretization optimality, without considering the time discretization optimality. Specifically, in [28] it was shown both theoretically and numerically that in the noDQ case the error bounds are suboptimal with respect to the ROM discretization error, whereas in the DQ case the error bounds are optimal. The time discretization optimality was ignored in [28].

In this paper, we prove (1.14). Specifically, we make three main contributions.

First, in the noDQ case, we prove that the POD error bound is suboptimal not only with respect to the ROM discretization (as shown in [28]) but also with respect

to the time discretization. Specifically, we show that the pointwise POD projection error assumption mentioned earlier can fail and the scaling of the error bound (1.12) with respect to the ROM discretization can degrade to

$$(1.15) \qquad \|e^n\|_{L^2}^2 = \mathcal{O}\left(\Delta t^{-1} \sum_{i=r+1}^{s} \lambda_i\right) + \mathcal{O}\left(\sum_{i=r+1}^{s} \lambda_i \|\nabla \varphi_i\|_{L^2}^2\right).$$

In particular, we construct two analytical examples, and we prove that they satisfy (1.15) in the noDQ case. We note that the bound (1.15) is a significant improvement over the bound (1.12) proved in [28], since the latter did not display the time discretization suboptimality.

Our second main contribution is that we prove new pointwise in time error bounds in the DQ case, and we do not require any of the assumptions used in [28] to establish similar pointwise bounds. All of these error bounds are optimal with respect to the time discretization. One key component of our analysis is that we prove that an assumption from [28, 29] concerning pointwise in time behavior of POD projection errors is automatically satisfied in the DQ case.

Our third main contribution is that we revisit the definition of ROM discretization error optimality, introduce a new stronger notion of optimality, and show that all of the pointwise in time error bounds in the DQ case are optimal in at least one sense. Both pointwise in time error bounds using the H_0^1 norm are optimal in the new stronger sense; the pointwise in time bounds using the L^2 norm can be optimal in either sense. We note that to prove the stronger optimality of the L^2 error bounds, we do need a uniform boundedness assumption of the type made in [28].

We emphasize that we do not attempt here to prove error bounds for the POD ROM when the parameter ν or the initial data are different from those used to generate the POD basis. As in [34] and many other POD numerical analysis works, our main goal here is to attempt to understand and begin to explain the approximation errors of POD ROMs for PDEs. The much more challenging case of analyzing errors in the POD ROM with variations in initial data and parameters is left to be explored elsewhere.

We also note that the analysis we perform is focused on a priori error estimates. For a posteriori error estimates extensive work has been conducted within the reduced basis community for parameterized PDEs (see, e.g., [17, 21, 45, 52, 53]). We also mention that a popular approach is to combine the reduced basis method in parameter space with POD in time [12]. Whether or not the results in this paper could be used to improve this approach is not explored within this paper.

DQs in applications. The focus of this paper is on the role played by DQs in the POD numerical analysis. We emphasize, however, that DQs are also widely used in practical applications.

One of the most important uses of DQs in practice is in *hyperreduction* methods for ROMs of nonlinear systems of the form $y' = f(t, y)$. Hyperreduction methods [57] significantly decrease the computational cost of the nonlinear ROM operator evaluations, which can be prohibitive in realistic applications. Popular hyperreduction methods (e.g., the empirical interpolation method (EIM) [3] and its discrete counterpart, the discrete empirical interpolation method (DEIM) [6]), use the nonlinear snapshots $f(t, y)$ to construct accurate approximations of the nonlinear ROM operators. As noted on page 48 in [6], since $f(t, y) = y'$ and $(y^{n+1} - y^n)/\Delta t \approx y'$, using nonlinear snapshots is similar to including the DQs. The DQs' connection to nonlinear snapshots was also used in [7] to develop the solution-based nonlinear subspace (SNS)

method as an efficient alternative to classical hyperreduction techniques. The SNS method was used in the reduced order modeling of the nonlinear diffusion equation and the parameterized quasi-one-dimensional Euler equation.

The DQs were explicitly used in various practical applications. For example, the DQs were utilized to develop data-driven ROMs for turbulent flows, in which the eddy viscosity field is a function of the time history of the velocity field (see section 3.3 in [22]). The DQs were also used in the reduced order modeling of the FitzHugh–Nagumo equations, which are used to model the dynamics of a spiking neuron (see section 4 in [33]). Furthermore, the DQs were employed to construct ROMs for the control of laser surface hardening [24], for feedback control of various PDEs [38], for partial integro-differential equations arising in financial applications [47], for subdiffusion equations [31], for convection-diffusion equations [59], for wave equations [20, 59], and for flow between offset cylinders and lid driven cavity flows [32].

In this paper, we use DQs with respect to time to obtain optimal pointwise in time error estimates. A different yet related approach was utilized in, e.g., [4, 30, 60], where DQs with respect to system parameters and initial conditions were used to improve the predictive capabilities of reduced basis methods [21, 43] for parameterized problems. In this setting, the noDQ case is referred to as the Lagrange approach, whereas the DQ case is referred to as the Hermite approach [30]. The error sensitivity with respect to parameters was investigated in, e.g., [25, 44].

The rest of the paper is organized as follows. In section 2, we describe the POD construction in the noDQ and DQ cases. In section 3, we give more detail about the previously described POD pointwise projection error assumption, show using examples that it can fail in the noDQ case, and prove that it is always satisfied in the DQ case. These results allow us to complete the POD ROM error analysis in section 4. For the first two main contributions, in section 5 we illustrate numerically the theoretical results. Specifically, for the heat equation (1.1) and both analytical examples, we show the following: (i) in the noDQ case, the error scales as in (1.15) (i.e., is suboptimal), and (ii) in the DQ case, the error scales according to the new error bounds. Finally, in section 6, we present our conclusions and future research directions.

2. Proper orthogonal decomposition. In this section we introduce two different approaches for constructing our reduced basis by using the POD [23, 54]. Suppose we have a collection of snapshots $U = \{u^n\}_{n=0}^N$ contained in a real Hilbert space $\mathcal{H}$. In the POD numerical analysis, a typical assumption (see, e.g., [34, 54]) is that each snapshot u^n is exactly equal to $u(t_n)$, where $u \in C([0,T]; \mathcal{H})$ and $t_n = n\Delta t$ for $n = 0, \ldots, N$ so that $t_0 = 0$, $t_N = T$, and $\Delta t = T/N$. For now, we only assume $T > 0$ is a fixed positive constant and we let $\Delta t = T/N$. We emphasize that T is fixed, but N is allowed to vary.

2.1. POD without difference quotients (noDQ case). We begin by examining the POD problem without DQs. In what follows, we denote this case the *noDQ case*. Given a fixed $r > 0$, the problem is to find a set of orthonormal basis functions $\{\varphi_i\}_{i=1}^r \subset \mathcal{H}$, called POD modes or POD basis functions, that optimally approximate the snapshots in the sense that the following error measure is minimized:

$$(2.1) \qquad E_r = \frac{1}{N+1} \sum_{n=0}^N \|u^n - P_r u^n\|_{\mathcal{H}}^2,$$

where $P_r : \mathcal{H} \to \mathcal{H}$ is the orthogonal projection onto $X^r = \text{span}\{\varphi_i\}_{i=1}^r$ given by

$$(2.2) \qquad P_r u = \sum_{i=1}^r (u, \varphi_i)_{\mathcal{H}} \varphi_i, \quad u \in \mathcal{H}.$$

One way to find a solution of this problem is to solve the eigenvalue problem

$$(2.3) \qquad K\mathbf{z}_i = \lambda_i \mathbf{z}_i \quad \text{for } i = 1, \dots, r,$$

where K is the snapshot correlation matrix with entries

$$(2.4) \qquad K_{mn} = \frac{1}{N+1} (u^m, u^n)_{\mathcal{H}}, \quad m, n = 0, \dots, N.$$

We order the eigenvalues $\{\lambda_i\}$ and corresponding orthonormal eigenvectors $\{\mathbf{z}_i\}$ so that $\lambda_1 \geq \lambda_2 \geq \lambda_{N+1} \geq 0$. The optimizing orthonormal set $\{\varphi_i\}_{i=1}^r \subset \mathcal{H}$ is given by

$$(2.5) \qquad \varphi_i = \lambda_i^{-1/2}(N+1)^{-1/2} \sum_{m=0}^N (\mathbf{z}_i)^m u^m, \quad i = 1, \dots, r,$$

where $(\mathbf{z}_i)^m$ is the mth entry of $\mathbf{z}_i$. Using these POD modes gives the optimal value for the approximation error:

$$(2.6) \qquad \frac{1}{N+1} \sum_{n=0}^N \|u^n - P_r u^n\|_{\mathcal{H}}^2 = \sum_{i>r} \lambda_i.$$

We note that the scaling factor $(N+1)^{-1}$ is important if one is interested in the solution of the optimization problem as more snapshots are collected, i.e., as Δt decreases or N increases. For certain choices of the scaling factor, the error measure E_r in (2.1) converges to a time integral or a constant multiple of a time integral, and the POD eigenvalues and POD modes also converge; see, e.g., [13, 18, 35, 49] for more information.

Different choices for the scaling factor in (2.1) have been used in the literature. We fix the scaling factor throughout this work to be $(N+1)^{-1}$ for simplicity. We note that since $\Delta t = T/N$, we have $(N+1)^{-1} = T_1^{-1}\Delta t$, where $T_1 = T + \Delta t$. Therefore, E_r in (2.1) is equal to the left Riemann sum approximation of the integral

$$\frac{1}{T_1} \int_0^{T_1} \|u(t) - P_r u(t)\|_{\mathcal{H}}^2 \, dt.$$

We note that the results in this work will hold for other scaling factors, as long as the scaling factor in question scales like a constant multiple of Δt.

Remark 2.1. One can also consider variable time steps and weights in the POD problem; we only consider a constant time step and single weight $(N+1)^{-1}$ for simplicity. Furthermore, one can use other quadrature rules, such as the midpoint rule or trapezoidal rule, to obtain appropriate weights for the POD problem.

In the following result, we give POD approximation errors in different norms and using other projections onto X^r. Similar results have been proved in multiple works (see, e.g., [28, 29, 39, 48, 50]), and our proof relies on techniques from these works. We note that this result can be obtained directly from the general results in the recent reference [39]; however, we include a proof to be complete. In this work, a bounded linear operator $\Pi : Z \to Z$ for a normed space Z is a projection onto $Z^r \subset Z$ if $\Pi^2 = \Pi$ and the range of Π equals Z^r. In this case, $\Pi z = z$ for any $z \in Z^r$.

LEMMA 2.2. *Let $X^r = \text{span}\{\varphi_i\}_{i=1}^r \subset \mathcal{H}$, let $P_r : \mathcal{H} \to \mathcal{H}$ be the orthogonal projection onto X^r as defined in* (2.2), *and let s be the number of positive POD eigenvalues for $U = \{u^n\}_{n=0}^N$. If W is a real Hilbert space with $U \subset W$ and $R_r : W \to W$ is a bounded linear projection onto X^r, then*

$$(2.7) \qquad \frac{1}{N+1} \sum_{n=0}^N \|u^n - P_r u^n\|_W^2 = \sum_{i=r+1}^s \lambda_i \|\varphi_i\|_W^2,$$

$$(2.8) \qquad \frac{1}{N+1} \sum_{n=0}^N \|u^n - R_r u^n\|_W^2 = \sum_{i=r+1}^s \lambda_i \|\varphi_i - R_r \varphi_i\|_W^2.$$

Proof. First, we note that (2.7) is a special case of (2.8) since $P_r \varphi_i = 0$ for $i > r$. Therefore, we only prove (2.8).

Next, by the POD approximation error formula (2.6), we have $u^n = P_s u^n$ for each n. If $r \geq s$, since R_r is a projection onto X^r we have $R_r u^n = R_r P_s u^n = P_s u^n = u^n$ and this proves the result. Therefore, assume $r < s$. Note by the definition of φ_i in (2.5), since $u^n \in W$ for each n we have $\varphi_i \in W$ for $i = 1, \ldots, r$. Therefore, $X^r \subset W$, and since the range of R_r equals X^r we know the W norm in (2.8) is well defined.

Now, using the definition of P_r in (2.2) gives

$$\frac{1}{N+1} \sum_{n=0}^N \|u^n - R_r u^n\|_W^2 = \frac{1}{N+1} \sum_{n=0}^N ((I - R_r) P_s u^n, (I - R_r) P_s u^n)_W$$

$$= \frac{1}{N+1} \sum_{n=0}^N \sum_{i,j=1}^s (u^n, \varphi_j)_{\mathcal{H}} (u^n, \varphi_i)_{\mathcal{H}} ((I - R_r)\varphi_j, (I - R_r)\varphi_i)_W,$$

where I is the identity operator. Next, take the $\mathcal{H}$ inner product of (2.5) with u^n and use the eigenvalue equations (2.3)–(2.4) to get

$$(u^n, \varphi_i)_{\mathcal{H}} = (N+1)^{1/2} \lambda_i^{1/2} (\mathbf{z}_i)^n.$$

Using this and also that $\{\mathbf{z}_i\}$ is orthonormal so that $\sum_{n=0}^N (\mathbf{z}_j)^n (\mathbf{z}_i)^n = \delta_{ij}$ gives

$$\frac{1}{N+1} \sum_{n=0}^N \|u^n - R_r u^n\|_W^2 = \sum_{i,j=1}^s (\lambda_i \lambda_j)^{1/2} \delta_{ij} ((I - R_r)\varphi_j, (I - R_r)\varphi_i)_W$$

$$= \sum_{i=1}^s \lambda_i \|(I - R_r)\varphi_i\|_W^2.$$

Since $\varphi_i \in X^r$ for $i = 1, \ldots, r$ and R_r is a projection onto X^r, we have $R_r \varphi_i = \varphi_i$ for $i = 1, \ldots, r$ and this proves the result. $\qquad \square$

2.2. POD with difference quotients (DQ case). In this section we consider a POD problem for the same snapshots as those in section 2.1, this time utilizing the DQs [34]: find an orthonormal set of basis functions $\{\varphi_i\}_{i=1}^r \subset \mathcal{H}$ minimizing the approximation error

$$(2.9) \qquad E_r^{\text{DQ}} = \frac{1}{2N+1} \sum_{n=0}^N \|u^n - P_r u^n\|_{\mathcal{H}}^2 + \frac{1}{2N+1} \sum_{n=1}^N \|\partial u^n - P_r \partial u^n\|_{\mathcal{H}}^2,$$

where the DQs $\{\partial u^n\}_{i=1}^N$ are defined by

$$(2.10) \qquad \partial u^n = \frac{u^n - u^{n-1}}{\Delta t}.$$

In what follows, we denote this case the *DQ case*.

Remark 2.3. One can give different weights to the snapshot and DQ approximation errors by replacing the second scaling factor $1/(2N+1)$ in (2.9) by a weighted fraction $\theta/(2N+1)$, where θ is a positive constant that is independent of N. The main results in this work can be modified to handle this case; however, we consider the unweighted case in (2.9) to simplify the presentation.

The solution to the minimization of the approximation error in (2.9) can be found by setting $v^n = u^n$ for $n = 0, \ldots, N$ and $v^{N+n} = \partial u^n$ for $n = 1, \ldots, N$. This yields a new collection of snapshots $U^{\mathrm{DQ}} = \{v^n\}_{n=0}^M$, where $M = 2N$. Proceeding as outlined in section 2.1 using the new collection $\{v^n\}_{n=0}^M$ in place of $\{u^n\}_{n=0}^N$ gives the solution of this different POD problem. We use $\{\lambda_i^{\mathrm{DQ}}\}$ to denote the POD eigenvalues for this POD problem; we use the same notation $\{\varphi_i\}_{i=1}^r$ for the POD basis functions. The optimal approximation error is given by

$$(2.11) \qquad \frac{1}{2N+1} \sum_{n=0}^N \|u^n - P_r u^n\|_{\mathcal{H}}^2 + \frac{1}{2N+1} \sum_{n=1}^N \|\partial u^n - P_r \partial u^n\|_{\mathcal{H}}^2 = \sum_{i>r} \lambda_i^{\mathrm{DQ}}.$$

Again, the choice of the scaling factor in the approximation error (2.9) is important if we consider the case where the amount of data increases, i.e., a finer time discretization is used so that Δt decreases and N increases. The DQs are used to approximate the time derivative of the data; therefore, for an appropriate choice of the scaling factor the approximation error in (2.9) contains approximations of time integrals involving both the data $u(t)$ and also the time derivative of the data $\partial_t u(t)$. For the DQ case, we use $(2N+1)^{-1}$ for the scaling factor throughout for simplicity.

As before, we give POD approximation errors in different norms and using other projections onto X^r.

LEMMA 2.4. *Let* $X^r = \mathrm{span}\{\varphi_i\}_{i=1}^r \subset \mathcal{H}$, *let* $P_r : \mathcal{H} \to \mathcal{H}$ *be the orthogonal projection onto* X^r *as defined in* (2.2), *and let* s *be the number of positive POD eigenvalues for the collection* $U^{\mathrm{DQ}} = \{v^n\}_{n=0}^{2N}$ *described above. If* W *is a real Hilbert space with* $U^{\mathrm{DQ}} \subset W$ *and* $R_r : W \to W$ *is a bounded linear projection onto* X^r, *then*

(2.12)

$$\frac{1}{2N+1} \left(\sum_{n=0}^N \|u^n - P_r u^n\|_W^2 + \sum_{n=1}^N \|\partial u^n - P_r \partial u^n\|_W^2 \right) = \sum_{i=r+1}^s \lambda_i^{\mathrm{DQ}} \|\varphi_i\|_W^2,$$

(2.13)

$$\frac{1}{2N+1} \left(\sum_{n=0}^N \|u^n - R_r u^n\|_W^2 + \sum_{n=1}^N \|\partial u^n - R_r \partial u^n\|_W^2 \right) = \sum_{i=r+1}^s \lambda_i^{\mathrm{DQ}} \|\varphi_i - R_r \varphi_i\|_W^2.$$

Proof. Apply Lemma 2.2 to the new collection of snapshots $\{v^n\}_{n=0}^M$ described above. □

Remark 2.5. In this section, we considered the DQs defined by (2.10). In practice the definition of the DQs will reflect the time discretization used to collect the snapshot data. For example, POD with central DQs is used for wave equations in [20, 59] and

fractional DQs are used for a subdiffusion problem in [31]. It is possible that the results of this paper can be extended to these and other definitions of the DQs, such as those arising from the backward differentiation formulas (BDF2, BDF3, etc.). We leave this to be considered elsewhere.

3. Pointwise projection error estimates. In the current literature on pointwise error bounds for the POD of parabolic problems several researchers make an assumption concerning the pointwise in time behavior of the POD projection errors [8, 9, 10, 19, 28, 29, 32, 41, 56, 58]. Roughly, the assumption says that the POD projection error at any time is of the same order as the total POD projection errors considered in section 2. Next, we formalize this assumption in Assumption 3.1, and then we discuss it for the noDQ case (section 3.1) and the DQ case (section 3.2).

We consider the POD of a collection of snapshots $U := \{u^n\}_{n=0}^N \subset \mathcal{H}$ and also $U \subset W$, as in section 2. Recall that $P_r : \mathcal{H} \to \mathcal{H}$ is the orthogonal projection onto the first r POD modes. For either the noDQ case or the DQ case, the pointwise POD projection error assumption is given as follows.

Assumption 3.1. There exists a constant C, depending on $T = N\Delta t$ only, such that the POD projection error satisfies

$$(3.1) \qquad \|u^n - P_r u^n\|_W^2 \leq C \sum_{i=r+1}^s \lambda_i \|\varphi_i\|_W^2 \quad \forall r = 1, \ldots, s \text{ and } n = 0, \ldots, N.$$

In section 3.1, we construct examples that show that this assumption can be violated in the noDQ case. In section 3.2, we show in Theorem 3.7 that this assumption is always satisfied in the DQ case.

Remark 3.1 (avoiding Assumption 3.1). We notice that Assumption 3.1 would follow directly from the POD approximation properties (2.7) (in the noDQ case) and (2.12) (in the DQ case) if we dropped the $1/(N+1)$ and $1/(2N+1)$ factors in the definitions (2.1) and (2.9) of the error measures E_r and E_r^{DQ}. In fact, when $\mathcal{H} = W = \mathbb{R}^m$, this approach is used in, e.g., [33]. We emphasize, however, that using this approach would increase by Δt^{-1} the magnitudes of the eigenvalues on the right-hand side of the POD approximation properties (2.7) and (2.12), which would yield suboptimal error estimates. Similar conclusions were reached in Remark 2.3 in [28] for the case $W = \mathcal{H}$.

Remark 3.2 (similar assumptions). For $W = \mathcal{H}$, Assumption 3.1 is Assumption 2.1 in [28] (in which the L^2 inner product should be replaced with the correct $\mathcal{H}$ inner product). A similar assumption (but for the L^2 projection of a continuous solution on X^r when $\mathcal{H} = L^2$) is made in Assumption 3.2 in [29]. No such assumption is made in [27], since Theorem 3.5 proves an estimate for the average error, not for the pointwise in time error. Finally, we note that Figure 4 in [28] provided numerical validation for Assumption 3.1 for the particular setting in [28] when $W = \mathcal{H}$.

3.1. Pointwise error estimates: noDQ case. First, we note that in general the scaling factor $N + 1$ is the worst-case scenario for the failure of Assumption 3.1. To see this, note that for any fixed k we have

$$\|u^k - P_r u^k\|_W^2 = (N+1)\frac{1}{N+1}\|u^k - P_r u^k\|_W^2$$

$$\tag{3.2} \leq (N+1)\left(\frac{1}{N+1}\sum_{i=0}^{N}\|u^i - P_r u^i\|_W^2\right)$$

$$\tag{3.3} = (N+1)\sum_{i=r+1}^{s}\lambda_i^{noDQ}\|\varphi_i\|_W^2,$$

where we used Lemma 2.2 to obtain (3.3). Note that for many collections of snapshots $\{u^k\}_{k=0}^{N}$ the inequality in (3.2) will be very conservative. Nevertheless, we show below that the above $N+1$ scaling is attained for a family of examples.

Assumption 3.1 says that the error at any particular index is not much larger than the other pointwise errors, or equivalently the inequality (3.2) is overly conservative. Therefore, Assumption 3.1 will be false if there is an index n such that the projection error at index n is much larger than the remaining pointwise errors, i.e.,

$$\tag{3.4} \|u^n - P_r u^n\|_W^2 \gg \|u^i - P_r u^i\|_W^2 \quad \forall i \neq n, \quad 0 \leq i \leq N.$$

Next, we provide a *family of counterexamples* to Assumption 3.1, i.e., a family of exact solutions (data) that yield POD bases that satisfy condition (3.4).

Let $\{\varphi_k\}_{k\geq 1}$ be an orthonormal set in a Hilbert space $\mathcal{H}$, with $\dim(\mathcal{H}) \geq N+1$, and let $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ be any sequence of positive numbers. Suppose the data $U = \{u^n\}_{n=0}^{N} \subset \mathcal{H}$ is given by

$$\tag{3.5} u^n = (N+1)^{1/2}\lambda_{n+1}^{1/2}\varphi_{n+1}, \quad n = 0, \ldots, N.$$

It can be checked that this data has POD eigenvalues $\{\lambda_k\}$ with corresponding POD modes $\{\varphi_k\}$.

Let W be a real Hilbert space with $U \subset W$. In Proposition 3.3, we show that Assumption 3.1 fails for the data above. Specifically, (3.6) shows that the assumption fails for the specific case of $r = N$ at index N. Furthermore, if the values $\{\lambda_k\}$ decay exponentially fast as in (3.7), then (3.8) shows that the assumption fails for any r at index r.

PROPOSITION 3.3. *Let the data $U = \{u^n\}_{n=0}^{N} \subset \mathcal{H}$ be given in (3.5) as described above. Then the POD pointwise projection error for u^N is given by*

$$\tag{3.6} \|u^N - P_N u^N\|_W^2 = (N+1)\lambda_{N+1}\|\varphi_{N+1}\|_W^2.$$

Also, for any fixed r if

$$\tag{3.7} \lambda_k = \beta\|\varphi_k\|_W^{-2}e^{-\gamma k}, \quad k > r,$$

for some positive constants β and γ, then

$$\tag{3.8} \|u^r - P_r u^r\|_W^2 \geq \frac{\min\{1,\gamma\}}{2}(N+1)\sum_{k=r+1}^{N+1}\lambda_k\|\varphi_k\|_W^2.$$

Remark 3.4. Note that for the second part of the result we still assume the POD eigenvalues in (3.7) are ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. Depending on the values of $\|\varphi_k\|_W$ and γ, the POD eigenvalues in (3.7) may not be ordered in this way. In such a case, the POD eigenvalues may need to be reordered in order to obtain a similar result. If $W = \mathcal{H}$ or if $\|\varphi_k\|_W$ increases slowly relative to $e^{-\gamma k}$, then the ordering $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ will automatically be satisfied.

Proof. Note that $P_r u^k = 0$ when $k \geq r$ and so

$$(3.9) \qquad \|u^k - P_r u^k\|_W^2 = (N+1)\lambda_{k+1}\|\varphi_{k+1}\|_W^2, \quad k \geq r.$$

Thus, (3.6) follows immediately from (3.9) with $k = N$.

Next, to prove (3.8), fix r and assume (3.7) holds. Then (3.9) with $k = r$ gives

$$(3.10) \qquad \|u^r - P_r u^r\|_W^2 = (N+1)\lambda_{r+1}\|\varphi_{r+1}\|_W^2.$$

We bound half of the right-hand side of (3.10) from below by a constant multiple of the remaining terms in the sum in (3.8). Note that the assumption (3.7) on the value of λ_{r+1} gives

$$(3.11) \qquad \frac{1}{2}\lambda_{r+1}\|\varphi_{r+1}\|_W^2 = \frac{\beta}{2}e^{-\gamma(r+1)}.$$

Next, we note that the exponential term on the right-hand side of (3.11) satisfies the following estimate:

$$(3.12) \qquad \frac{1}{\gamma}e^{-\gamma(r+1)} \geq \frac{1}{\gamma}\left(e^{-\gamma(r+1)} - e^{-\gamma(N+1)}\right) = \int_{r+1}^{N+1} e^{-\gamma x}dx \geq \sum_{k=r+2}^{N+1} e^{-\gamma k}.$$

Using (3.7), (3.11), and (3.12), we obtain

$$(3.13)$$
$$\frac{1}{2}(N+1)\lambda_{r+1}\|\varphi_{r+1}\|_W^2 \geq \frac{\gamma\beta}{2}(N+1)\sum_{k=r+2}^{N+1} e^{-\gamma k} = \frac{\gamma}{2}(N+1)\sum_{k=r+2}^{N+1} \lambda_k\|\varphi_k\|_W^2.$$

Using (3.10) and (3.13), we get

$$\|u^r - P_r u^r\|_W^2 \geq \frac{1}{2}(N+1)\lambda_{r+1}\|\varphi_{r+1}\|_W^2 + \frac{\gamma}{2}(N+1)\sum_{k=r+2}^{N+1} \lambda_k\|\varphi_k\|_W^2$$

$$(3.14) \qquad\qquad \geq \frac{\min\{1,\gamma\}}{2}(N+1)\sum_{k=r+1}^{N+1} \lambda_k\|\varphi_k\|_W^2,$$

which proves (3.8). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Proposition 3.3 yields a family of counterexamples to Assumption 3.1. Next, we consider two counterexamples that we investigate numerically in section 5.

3.1.1. Counterexample 1. To construct the first counterexample to Assumption 3.1 (which we denote counterexample 1), we follow the theoretical setting in this section and construct a family of ROM basis functions that satisfy (3.5). Specifically, we consider an orthonormal set $\{\varphi_n\}_{n=0}^N$ in $\mathcal{H} = L^2(0,1)$ given by

$$(3.15) \qquad \varphi_{n+1}(x) := 2^{1/2}\sin((k\,t_n+1)\pi\,x),$$

where k is a positive integer, $x \in [0,1]$, and $t_n = n\,\Delta t$ is chosen such that $k\,t_n \in \mathbb{N}\,\forall\,n \in \mathbb{N}$. Next, we choose the eigenvalues

$$(3.16) \qquad \lambda_1 = \lambda_2 = \cdots = \lambda_{N+1} = \frac{1}{2(N+1)},$$

which satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N+1} > 0$. Finally, choosing the analytical solution

$$(3.17) \qquad u_{\text{counterexample 1}}(x, t) = \sin((k\,t + 1)\pi\,x)$$

and the corresponding forcing term

$$(3.18) \qquad f = (k\,\pi\,x)\cos((k\,t + 1)\pi\,x) + \nu\pi^2(kt + 1)^2 \sin((k\,t + 1)\pi\,x)$$

yields the data $U = \{u^n\}_{n=0}^{N}$ that satisfies (3.5).

In section 5, we investigate numerically counterexample 1 given by the analytical solution (3.17).

Remark 3.5. Equation (3.5) (see also the comment below Assumption A.1 in [41]) shows that the ROM basis functions are scaled versions of the snapshots. For counterexample 1, this scaling is illustrated in (3.15) and (3.17).

3.1.2. Counterexample 2. To construct the second counterexample to Assumption 3.1 (which we denote counterexample 2), we construct a family of ROM basis functions that satisfy both (3.5) and (3.7) in Proposition 3.3. Specifically, we consider the same orthonormal set $\{\varphi_n\}_{n=0}^{N}$ in $\mathcal{H} = L^2(0, 1)$ given in (3.15) above, where again k is a positive integer, $x \in [0, 1]$, and $t_n = n\,\Delta t$ is chosen such that $k\,t_n \in \mathbb{N}$, $\forall n \in \mathbb{N}$. Next, for positive constants α, δ, and ρ, with $\delta = \rho\Delta t$, we choose exponentially decaying eigenvalues as in (3.7):

$$\lambda_{n+1} = \beta e^{-\gamma(n+1)},$$
$$\beta = \frac{1}{4\delta(N+1)}e^{-\alpha+\alpha\delta^{-1}\Delta t} = \frac{1}{4\rho T_1}e^{-\alpha+\alpha\rho^{-1}},$$
$$\gamma = \alpha\delta^{-1}\Delta t = \alpha\rho^{-1},$$

where $T_1 = T + \Delta t$. Finally, it can be checked that choosing the analytical solution

$$(3.19) \qquad u_{\text{counterexample 2}}(x, t) = \frac{1}{\sqrt{2\delta}}\left(e^{-\alpha(1+t/\delta)}\right)^{1/2}\sin((kt + 1)\pi x)$$

and the corresponding forcing term

$$(3.20) \qquad \begin{aligned} f = \frac{1}{\sqrt{2\delta}}\left(e^{-\alpha(1+t/\delta)}\right)^{1/2}&\left[\frac{-\alpha}{2\delta}\sin((kt + 1)\pi x) + (k\pi x)\cos((kt + 1)\pi x)\right.\\ &\left.+ \nu\pi^2(kt + 1)^2 \sin((kt + 1)\pi x)\right]\end{aligned}$$

yields the data $U = \{u^n\}_{n=0}^{N}$ that satisfies (3.5), which shows that, in counterexample 2, the ROM basis functions are scaled versions of the snapshots.

In section 5, we investigate numerically counterexample 2 given by the analytical solution (3.19).

3.2. POD pointwise error estimates: DQ case. We now give one of the main results of this paper. In Theorem 3.7, we show that Assumption 3.1 is always satisfied in the DQ case. This will allow us to prove in section 4 optimal pointwise in time ROM error bounds in the DQ case. In particular, Theorem 3.7 will show that the assumptions similar to Assumption 3.1 that have been made in, e.g., [28] are unnecessary for obtaining optimal error bounds in the DQ case.

In continuous time, it is well known that the magnitude of a function $z \in H^1(0, T)$ at any point in time is bounded above by a constant multiple of the $H^1(0, T)$ norm

of z. The constant in the bound only depends on T, and there is also a similar inequality that holds for functions taking values in a Banach space Z (see, e.g., [11, section 5.9.2, Theorem 2(iii)], p. 302). Below, we establish a discrete time analogue of this Sobolev embedding $H^1(0, T; Z) \hookrightarrow C([0, T]; Z)$, where the DQs replace the time derivative in the $H^1(0, T; Z)$ norm. This lemma will allow us to directly establish POD pointwise projection error bounds in Theorem 3.7, which shows that Assumption 3.1 is automatically satisfied in the DQ case.

LEMMA 3.6 (discrete time Sobolev inequality). *Let $T > 0$, Z be a normed space, $\{z^n\}_{n=0}^{N} \subset Z$, and $\Delta t = T/N$. Then*

$$\max_{0 \leq k \leq N} \|z^k\|_Z^2 \leq C \left(\frac{1}{2N+1} \sum_{n=0}^{N} \|z^n\|_Z^2 + \frac{1}{2N+1} \sum_{n=1}^{N} \|\partial z^n\|_Z^2 \right),$$

where $C = 6 \max\{1, T^2\}$ and $\partial z^n = (z^n - z^{n-1})/\Delta t$ for $n = 1, \dots, N$.

Proof. For each k, ℓ with $N \geq k > \ell \geq 0$, we have $z^k - z^\ell = \Delta t \sum_{n=\ell+1}^{k} \partial z^n$. This gives

(3.21)

$$\|z^k\|_Z \leq \|z^\ell\|_Z + \sum_{n=1}^{N} \Delta t^{1/2}(\Delta t^{1/2}\|\partial z^n\|_Z) \leq \|z^\ell\|_Z + T^{1/2} \left(\sum_{n=1}^{N} \Delta t \|\partial z^n\|_Z^2 \right)^{1/2},$$

where we used $\sum_{n=1}^{N} \Delta t = N\Delta t = T$. This inequality is also clearly true for $k = \ell$, and a similar argument shows that this inequality also holds for $0 \leq k < \ell \leq N$.

Now we choose ℓ so that

(3.22)
$$\|z^\ell\|_Z = \min_{0 \leq n \leq N} \|z^n\|_Z.$$

We know such an ℓ must exist since N is finite. Then

$$\|z^\ell\|_Z = \frac{1}{N+1}(N+1)\|z^\ell\|_Z = \frac{1}{N+1} \sum_{n=0}^{N} \|z^\ell\|_Z$$

$$\leq \frac{1}{T} \sum_{n=0}^{N} \Delta t \|z^n\|_Z \leq T^{-1/2} \left(\sum_{n=0}^{N} \Delta t \|z^n\|_Z^2 \right)^{1/2},$$

where we used (3.22), $1/(N+1) < 1/N = T^{-1}\Delta t$, $\sum_{n=1}^{N} \Delta t = N\Delta t = T$, and the Cauchy–Schwarz inequality. Using this inequality with (3.21) yields

(3.23) $$\|z^k\|_Z \leq T^{-1/2} \left(\sum_{n=0}^{N} \Delta t \|z^n\|_Z^2 \right)^{1/2} + T^{1/2} \left(\sum_{n=1}^{N} \Delta t \|\partial z^n\|_Z^2 \right)^{1/2}.$$

Squaring both sides, and using the inequalities $(a + b)^2 \leq 2(a^2 + b^2)$ and $\Delta t = (2T + \Delta t)/(2N + 1) \leq 3T/(2N + 1)$, we obtain the result. $\square$

THEOREM 3.7. *Let $X^r = \mathrm{span}\{\varphi_i\}_{i=1}^{r} \subset \mathcal{H}$, let $P_r : \mathcal{H} \to \mathcal{H}$ be the orthogonal projection onto X^r as defined in (2.2), and let s be the number of positive POD eigenvalues for U^{DQ}. If W is a real Hilbert space with $U^{\mathrm{DQ}} \subset W$ and $R_r : W \to W$ is a bounded linear projection onto X^r, then*

$$(3.24a) \qquad \max_{0 \le k \le N} \left\| u^k - P_r u^k \right\|_{\mathcal{H}}^2 \le C \sum_{i=r+1}^{s} \lambda_i^{\mathrm{DQ}},$$

$$(3.24b) \qquad \max_{0 \le k \le N} \left\| u^k - P_r u^k \right\|_W^2 \le C \sum_{i=r+1}^{s} \lambda_i^{\mathrm{DQ}} \|\varphi_i\|_W^2,$$

$$(3.24c) \qquad \max_{0 \le k \le N} \left\| u^k - R_r u^k \right\|_W^2 \le C \sum_{i=r+1}^{s} \lambda_i^{\mathrm{DQ}} \|\varphi_i - R_r \varphi_i\|_W^2,$$

where $C = 6 \max\{1, T^2\}$.

Proof. First, note that (3.24a) follows from (3.24b) with $W = \mathcal{H}$ since $\|\varphi_i\|_{\mathcal{H}} = 1 \ \forall i$. Also, (3.24b) follows from (3.24c) since $P_r \varphi_i = 0$ for $i > r$. Therefore, we only prove (3.24c).

Set $Z = W$ and $z^n = u^n - R_r u^n$ for each n. Using Lemma 3.6, $\partial z^n = \partial u^n - R_r \partial u^n$ for each n, and Lemma 2.4 gives the result. $\qquad \square$

4. Pointwise error estimates: DQ case. In this section, we prove pointwise in time error estimates for the heat equation and discuss the time and ROM discretization optimality of these estimates. In section 4.1, we prove the pointwise in time error estimates using CN time stepping in the DQ case (see section 2.2). In section 4.2, we consider three definitions of optimality for the ROM discretization error and classify the optimality types of each pointwise error estimate in section 4.1. We show that all of the error estimates are optimal in some sense, although in some cases we need to assume that various POD projection uniform boundedness conditions are satisfied. We also briefly discuss error estimates and optimality for the noDQ case; see Remarks 4.2, 4.4, and 4.9. Below, we consider the DQ case unless explicitly mentioned otherwise.

We begin by establishing notation and definitions and giving preliminary results that will be used in the ensuing analysis. We let $\Omega \in \mathbb{R}^d, d = 2, 3$, be a regular open domain with Lipschitz continuous boundary Ω and denote by $(\cdot, \cdot)_{L^2}$ and $\| \cdot \|_{L^2}$ the L^2 inner product and norm, respectively. We define the function space $X = H_0^1(\Omega)$ as

$$X := H_0^1(\Omega)^d = \{ v \in H^1(\Omega)^d : v|_\Gamma = 0 \}.$$

With the inner product $(u, v)_{H_0^1} = (\nabla u, \nabla v)_{L^2}$, the space $X = H_0^1(\Omega)$ is a Hilbert space.

For simplicity, we will only consider the heat equation (1.1). We take $u(\cdot, t) \in X$, $t \in [0, T]$ to be the weak solution of the weak formulation of the heat equation with homogeneous Dirichlet boundary conditions:

$$(4.1) \qquad (\partial_t u, v)_{L^2} + \nu (\nabla u, \nabla v)_{L^2} = (f, v)_{L^2} \quad \forall v \in X.$$

Replacing the unknown u with u_r in the heat equation (4.1), using the Galerkin method, projecting the resulting equations onto a space $X^r \subset X$, and discretizing in time using CN, one obtains the standard CN POD-G-ROM for the heat equation

$$(4.2) \qquad (\partial u_r^{n+1}, v_r)_{L^2} + \nu (\nabla u_r^{n+1/2}, \nabla v_r)_{L^2} = (f^{n+1/2}, v_r)_{L^2} \quad \forall v_r \in X^r,$$

where $\partial u_r^{n+1} = (u_r^{n+1} - u_r^n)/\Delta t$. Also, here and below we use the notation $z^{n+1/2}$ for any discrete or continuous time function z to denote the average

$$z^{n+1/2} := \frac{1}{2} \left(z^{n+1} + z^n \right).$$

Note that, for continuous time functions, we do *not* use $z^{n+1/2}$ to denote $z(t_n + \Delta t/2)$.

Remark 4.1. An alternative CN approach to the time discretization is to replace $f^{n+1/2}$ in (4.2) with $f(t_n + \Delta t/2)$. The results in this section also hold for this case.

We now prove error estimates for the error $u^{n+1} - u_r^{n+1}$, where $u^{n+1} := u(t_{n+1})$ is the solution of the weak formulation of the heat equation (4.1), and u_r^{n+1} is the solution of the CN POD-G-ROM (4.2). For clarity of presentation, we only consider the error components corresponding to the POD truncation and time discretization, i.e., we ignore the spatial discretization (e.g., FE) error. We start by noting that the weak solution of the heat equation evaluated at time $t = t_n + \Delta t/2$ satisfies

$$(4.3) \quad \left(\partial u^{n+1}, v_r\right)_{L^2} + \nu(\nabla u^{n+1/2}, \nabla v_r)_{L^2} = (f^{n+1/2}, v_r)_{L^2} + \tau_n(v_r) \quad \forall v_r \in X^r,$$

where $\partial u^{n+1} = (u^{n+1} - u^n)/\Delta t$, and, after integrating by parts, the consistency error is given by

$$(4.4) \quad \begin{aligned} \tau_n(v) &:= \left(\partial u^{n+1} - \partial_t u(t_n + \Delta t/2), v\right)_{L^2} + \nu \left(\Delta(u(t_n + \Delta t/2) - u^{n+1/2}), v\right)_{L^2} \\ &\quad + \left(f(t_n + \Delta t/2) - f^{n+1/2}, v\right)_{L^2}. \end{aligned}$$

We assume that the solution u and the forcing f are smooth enough so that $\tau_n(v)$ is well defined for any $v \in X$. We provide a more precise regularity assumption below.

The error is split into two parts:

$$(4.5) \quad e^{n+1} = u^{n+1} - u_r^{n+1} = (u^{n+1} - w_r^{n+1}) - (u_r^{n+1} - w_r^{n+1}) = \eta^{n+1} - \phi_r^{n+1},$$

where w_r^{n+1} is a proper projection of u^{n+1} on X^r, $\eta^{n+1} := u^{n+1} - w_r^{n+1}$, and $\phi_r^{n+1} = u_r^{n+1} - w_r^{n+1}$. Subtracting (4.2) from (4.3) then yields

$$(4.6) \quad \begin{aligned} (\partial \phi_r^{n+1}, v_r)_{L^2} + \nu(\nabla \phi_r^{n+1/2}, \nabla v_r)_{L^2} &= (\partial \eta^{n+1}, v_r)_{L^2} + \nu(\nabla \eta^{n+1/2}, \nabla v_r)_{L^2} \\ &\quad - \tau_n(v_r) \quad \forall v_r \in X^r. \end{aligned}$$

The standard approach used to prove error estimates in this case is to use the Ritz projection [2, 27, 29, 34, 35, 46]. This is also the standard approach in the FE context [15, 37, 51, 55]. Thus, for the ensuing analysis we choose $w_r := R_r(u)$ in (4.5), where $R_r(u)$ is the Ritz projection of u on X^r:

$$(4.7) \quad (\nabla(u - R_r(u)), \nabla v_r)_{L^2} = 0 \quad \forall v_r \in X^r.$$

We then denote $\eta_{Ritz} := u - R_r(u)$. Using the Ritz projection, (4.6) then becomes

$$(4.8) \quad (\partial \phi_r^{n+1}, v_r)_{L^2} + \nu(\nabla \phi_r^{n+1/2}, \nabla v_r)_{L^2} = (\partial \eta_{Ritz}^{n+1}, v_r)_{L^2} - \tau_n(v_r) \quad \forall v_r \in X^r,$$

where we have used the fact that $(\nabla \eta_{Ritz}^{n+1/2}, \nabla v_r)_{L^2} = 0$ by (4.7).

Remark 4.2. In the noDQ case (see section 2.1), a different approach is typically used to prove error estimates; see, e.g., [5, 28, 29, 50]. Instead of the Ritz projection, in the noDQ case we use the L^2 projection $\Pi_r^{L^2}$ and take $w_r^{n+1} = \Pi_r^{L^2} u^{n+1}$. The term $\nu(\nabla \eta^{n+1/2}, \nabla v_r)_{L^2}$ in (4.6) no longer vanishes; instead, the DQ projection error term is eliminated, i.e., $(\partial \eta^{n+1}, v_r)_{L^2} = 0$ in (4.6). However, as explained in Remark 4.4, the resulting pointwise error estimates are suboptimal.

For the POD basis construction, we must specify a Hilbert space $\mathcal{H}$. For this problem, two natural Hilbert spaces that are often used are $\mathcal{H} = L^2(\Omega)$ or $\mathcal{H} = X = H_0^1(\Omega)$. Let X^r be the span of the first r POD modes for the data set containing

the snapshots $\{u^n\}_{n=0}^N$ and the snapshot DQs $\{\partial u^n\}_{n=1}^N$. We can use Lemma 2.4 and Theorem 3.7 to obtain POD approximation error results with either $W = L^2(\Omega)$ or $W = H_0^1(\Omega)$. We note that in the case $\mathcal{H} = H_0^1(\Omega)$, the standard orthogonal POD projection P_r is exactly equal to the Ritz projection R_r.

We emphasize that, in this section, we use the exact solution of the heat equation for the POD basis construction in order to focus on the POD and time discretization errors. Exact solution data was used in this way by Kunisch and Volkwein in their original POD numerical analysis work [34] and also by many other researchers in subsequent works.

4.1. Error estimates. We give multiple error bounds for the solution when both the L^2 and H_0^1 POD bases are used. Specifically, we first provide a pointwise in time error bound for the L^2 norm of the solution, and an error bound for the solution norm (a discrete time analogue of the $L^2(0, T; H_0^1(\Omega))$ norm) that includes the L^2 norm of the solution at the final time step. Then, we prove a pointwise in time error bound for the H_0^1 norm of the solution.

We assume that the solution u of the heat equation (1.1) and the forcing f satisfy the regularity condition

$$(4.9) \qquad u_{ttt}, \ \Delta u_{tt}, \ f_{tt} \in L^2(0, T; L^2(\Omega)).$$

We also define the regularity constants

$$I_{n,1}(u,f) := \|u_{ttt}\|_{L^2(t_n,t_{n+1};L^2)} + \|\Delta u_{tt}\|_{L^2(t_n,t_{n+1};L^2)} + \|f_{tt}\|_{L^2(t_n,t_{n+1};L^2)},$$

$$(4.10) \qquad I_n(u,f) := \|u_{ttt}\|_{L^2(t_n,t_{n+1};L^2)}^2 + \|\Delta u_{tt}\|_{L^2(t_n,t_{n+1};L^2)}^2 + \|f_{tt}\|_{L^2(t_n,t_{n+1};L^2)}^2,$$

$$I(u,f) := \|u_{ttt}\|_{L^2(0,T;L^2)}^2 + \|\Delta u_{tt}\|_{L^2(0,T;L^2)}^2 + \|f_{tt}\|_{L^2(0,T;L^2)}^2.$$

As mentioned above, for all of the results below we assume that X^r is the span of the first r POD modes for the data set containing the snapshots $\{u^n\}_{n=0}^N$ of the exact solution and the snapshot DQs $\{\partial u^n\}_{n=1}^N$. Furthermore, as pointed out in the introduction, we prove POD error bounds when the parameter ν and the initial data are the same as those used to generate the POD basis.

LEMMA 4.3. *Consider the CN POD-G-ROM scheme* (4.2). *If* (4.9) *is satisfied, then the following error bounds hold when the L^2 POD basis is used,*

$$(4.11) \qquad \max_{1 \le k \le N} \|e^k\|_{L^2}^2 \le C \left(\sum_{i=r+1}^s \lambda_i^{DQ} \|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \|\phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f) \right),$$

$$(4.12)$$
$$\|e^N\|_{L^2}^2 + \Delta t \sum_{n=0}^{N-1} \|\nabla e^{n+1/2}\|_{L^2}^2 \le C \left(\sum_{i=r+1}^s \lambda_i^{DQ} (\|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 \right.$$
$$\left. + \|\nabla(\varphi_i - R_r(\varphi_i))\|_{L^2}^2) + \|\phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f) \right),$$

and the following error bounds hold when the H_0^1 POD basis is used,

$$(4.13) \qquad \max_{1 \le k \le N} \|e^k\|_{L^2}^2 \le C \left(\sum_{i=r+1}^s \lambda_i^{DQ} \|\varphi_i\|_{L^2}^2 + \|\phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f) \right),$$

(4.14)

$$\|e^N\|_{L^2}^2 + \Delta t \sum_{n=0}^{N-1} \|\nabla e^{n+1/2}\|_{L^2}^2 \leq C\left(\sum_{i=r+1}^s (1 + \|\varphi_i\|_{L^2}^2)\lambda_i^{DQ} + \|\phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f)\right).$$

Proof. We let $v_r := \phi_r^{n+1/2}$ in (4.8) and use $2\Delta t\,(\partial\phi_r^{n+1}, \phi_r^{n+1/2})_{L^2} = \|\phi_r^{n+1}\|_{L^2}^2 - \|\phi_r^n\|_{L^2}^2$ to obtain

(4.15)
$$\|\phi_r^{n+1}\|_{L^2}^2 - \|\phi_r^n\|_{L^2}^2 + 2\nu\Delta t\|\nabla\phi_r^{n+1/2}\|_{L^2}^2 = 2\Delta t[(\partial\eta_{Ritz}^{n+1}, \phi_r^{n+1/2})_{L^2} \\ - \tau_n(\phi_r^{n+1/2})].$$

Next, we use the Cauchy–Schwarz inequality, the Poincaré inequality $\|\phi_r^{n+1/2}\|_{L^2} \leq C\|\nabla\phi_r^{n+1/2}\|_{L^2}$, and Taylor's theorem[1] to bound the right-hand side and obtain

(4.16)
$$\|\phi_r^{n+1}\|_{L^2}^2 - \|\phi_r^n\|_{L^2}^2 + 2\nu\Delta t\|\nabla\phi_r^{n+1/2}\|_{L^2}^2 \\ \leq C\Delta t\big(\|\partial\eta_{Ritz}^{n+1}\|_{L^2} + \Delta t^{3/2}I_{n,1}(u,f)\big)\|\nabla\phi_r^{n+1/2}\|_{L^2}.$$

Applying Young's inequality and using $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ yields

(4.17)

$$\|\phi_r^{n+1}\|_{L^2}^2 - \|\phi_r^n\|_{L^2}^2 + 2\nu\Delta t\,\|\nabla\phi_r^{n+1/2}\|_{L^2}^2 \leq \Big(C\Delta t\,\big\|\partial\eta_{Ritz}^{n+1}\big\|_{L^2}^2 + C\Delta t^4 I_n(u,f) \\ + \nu\Delta t\,\|\nabla\phi_r^{n+1/2}\|_{L^2}^2\Big).$$

Now, summing from $n=0$ to $k-1$ gives
(4.18)
$$\|\phi_r^k\|_{L^2}^2 + \nu\sum_{n=0}^{k-1}\Delta t\,\|\nabla\phi_r^{n+1/2}\|_{L^2}^2 \leq C\left(\sum_{n=0}^{k-1}\Delta t\,\big\|\partial\eta_{Ritz}^{n+1}\big\|_{L^2}^2 + \Delta t^4 I(u,f) + \|\phi_r^0\|_{L^2}^2\right).$$

By the triangle inequality we have $\|e^k\|_{L^2}^2 \leq 2(\|\eta_{Ritz}^k\|_{L^2}^2 + \|\phi_r^k\|_{L^2}^2)$. Applying this inequality, rearranging terms, dropping an unnecessary term, and taking a maximum among constants it then follows from (4.18) that

(4.19) $$\|e^k\|_{L^2}^2 \leq C\left(\Delta t\sum_{n=1}^N \|\partial\eta_{Ritz}^n\|_{L^2}^2 + \|\eta_{Ritz}^k\|_{L^2}^2 + \|\phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f)\right).$$

The pointwise in time estimates (4.11) and (4.13) then follow from applying Lemma 2.4 and Theorem 3.7 with $W = L^2(\Omega)$ and $\eta_{Ritz}^k = u^k - R_r u^k$, where R_r is the Ritz projection (which also equals P_r for the H_0^1 POD basis), and using $\Delta t(2N+1) = (2+1/N)T \leq 3T$.

The error bounds (4.12) and (4.14) in the solution norm follow by taking $k = N$ in (4.18) and proceeding similarly. $\square$

Remark 4.4. We briefly provide one pointwise in time error estimate for the noDQ case with the L^2 POD basis; other pointwise estimates can be obtained using similar ideas. In the noDQ case, to obtain a pointwise in time L^2 error estimate one can

---

[1]For more details, see, e.g., [37, Lemma 26, p. 166] or [51, pp. 16–17].

proceed in a similar fashion to the above proof using the L^2 projection instead of the Ritz projection, as discussed in Remark 4.2. The error estimate (4.20) can be obtained using Lemma 2.2 with $\mathcal{H} = L^2(\Omega)$ and $W = H_0^1(\Omega)$, and the worst-case pointwise projection error bound (3.3):

$$
(4.20) \quad \max_{1 \le k \le N} \|e^k\|_{L^2}^2 \le C\Bigg((N+1) \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} + \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} \|\nabla \varphi_i\|_{L^2}^2
$$
$$
+ \|\phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f) \Bigg).
$$

If Assumption 3.1 is satisfied, then the $(N+1)$ scaling factor can be removed.

We emphasize that the error estimate (4.20) is suboptimal; see Remark 4.9 below for precise optimality definitions. First, the estimate is suboptimal with respect to the time discretization error because of the extra factor $(N+1) = (T\Delta t^{-1} + 1)$. Second, the estimate is suboptimal with respect to the POD projection error because of the second term on the right-hand side, which contains $\|\nabla \varphi_i\|_{L^2}^2$ instead of $\|\varphi_i\|_{L^2}^2$. This is a consequence of using the L^2 projection instead of the classical Ritz projection (see Remark 4.2). As explained in [28], using the L^2 projection eliminates the need to use the DQs but yields suboptimal estimates with respect to the POD projection error. Thus, even if Assumption 3.1 is satisfied and the $(N+1)$ scaling factor can be removed, the error estimate (4.20) is still suboptimal. If the H_0^1 POD basis is used instead, the resulting error estimate is also suboptimal, even if Assumption 3.1 is satisfied; the details are similar.

Next, we prove a pointwise in time error bound in the H_0^1 norm.

LEMMA 4.5. *Consider the CN POD-G-ROM scheme* (4.2). *If* (4.9) *is satisfied, then the following error bound holds when the L^2 POD basis is used,*

$$
(4.21) \quad \max_{1 \le k \le N} \|\nabla e^k\|_{L^2}^2 \le C \left(\sum_{i=r+1}^{s} \lambda_i^{DQ} \left(\|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \|\nabla(\varphi_i - R_r(\varphi_i))\|_{L^2}^2 \right) \right.
$$
$$
\left. + \|\nabla \phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f) \right),
$$

and the following error bound holds when the H_0^1 POD basis is used,

$$
(4.22) \quad \max_{1 \le k \le N} \|\nabla e^k\|_{L^2}^2 \le C \left(\sum_{i=r+1}^{s} \lambda_i^{DQ}(1 + \|\varphi_i\|_{L^2}^2) + \|\nabla \phi_r^0\|_{L^2}^2 + \Delta t^4 I(u,f) \right).
$$

Proof. We let $v_r := \partial \phi_r^{n+1}$ in (4.8):

$$
(4.23)
$$
$$
\|\partial \phi_r^{n+1}\|_{L^2}^2 + \frac{\nu}{2\Delta t}(\|\nabla \phi_r^{n+1}\|_{L^2}^2 - \|\nabla \phi_r^n\|_{L^2}^2) = (\partial \eta_{Ritz}^{n+1}, \partial \phi_r^{n+1})_{L^2} - \tau_n(\partial \phi_r^{n+1}).
$$

Applying Cauchy–Schwarz and Young's inequalities along with Taylor's theorem on the right-hand side of (4.23), we get

$$
(4.24)
$$
$$
\nu(\|\nabla \phi_r^{n+1}\|_{L^2}^2 - \|\nabla \phi_r^n\|_{L^2}^2) + 2\Delta t \|\partial \phi_r^{n+1}\|_{L^2}^2 \le \Delta t \|\partial \eta_{Ritz}^{n+1}\|_{L^2}^2 + \frac{3}{2}\Delta t \|\partial \phi_r^{n+1}\|_{L^2}^2
$$
$$
+ C\Delta t^4 I_n(u,f).
$$

Next, we sum from $n = 0$ to $n = k - 1$ and drop an unnecessary term:

$$\|\nabla \phi_r^k\|_{L^2}^2 \leq \frac{1}{\nu} \sum_{n=0}^{N-1} \Delta t \, \|\partial \eta_{Ritz}^{n+1}\|_{L^2}^2 + C \Delta t^4 I(u, f) + \|\nabla \phi_r^0\|_{L^2}^2.$$

Now we use $\|\nabla e^k\|_{L^2}^2 \leq 2(\|\nabla \eta_{Ritz}^k\|_{L^2}^2 + \|\nabla \phi_r^k\|_{L^2}^2)$ to obtain

$$\|\nabla e^k\|_{L^2}^2 \leq C \left(\sum_{n=0}^{N-1} \Delta t \, \|\partial \eta_{Ritz}^{n+1}\|_{L^2}^2 + \|\nabla \eta_{Ritz}^k\|_{L^2}^2 + \Delta t^4 I(u, f) + \|\nabla \phi_r^0\|_{L^2}^2 \right).$$

We use Lemma 2.4, Theorem 3.7, and $\Delta t(2N + 1) = (2 + 1/N)T \leq 3T$ to complete the proof. $\qquad\square$

4.2. Optimality of pointwise ROM discretization errors. Next, we discuss three different definitions of optimality for pointwise in time ROM discretization errors. Again, we assume that we are in the DQ case throughout, although we do briefly discuss the noDQ case in Remark 4.9 below. We classify the optimality type of each pointwise in time error bound for the DQ case from section 4.1.

The optimality type of a pointwise error bound depends on both the space $\mathcal{H}$ for the POD basis and the space W for the pointwise error norm. In section 4.1 we considered four possibilities: we used $\mathcal{H} = L^2$ or $\mathcal{H} = H_0^1$ for the POD basis, and we used $W = L^2$ or $W = H_0^1$ for the error norm. Below, we let $\mathcal{H}$ and W be any real Hilbert spaces, we consider the DQ case, and we let $e^k = u^k - u_r^k$ be the ROM error for $k = 0, \ldots, N$. For the discretization, we assume that if certain conditions are satisfied, then there exists a constant C so that the following pointwise error bound holds:

$$(4.25) \qquad \max_{1 \leq k \leq N} \|e^k\|_W^2 \leq C \left(\Lambda_r + \Lambda_r^0 + \zeta(\Delta t) + \xi(h) \right),$$

where

- Λ_r is the ROM discretization error and depends only on r, the POD eigenvalues, and the POD modes;
- Λ_r^0 is the ROM discretization error for the initial condition only and depends only on r, the POD eigenvalues, and the POD modes;
- $\zeta(\Delta t)$ is an *optimal* time discretization error; and
- $\xi(h)$ is an *optimal* spatial discretization error.

We automatically consider the discretization error suboptimal if either the time or the space discretization errors are suboptimal; therefore, we assume those errors are optimal here and focus on the ROM discretization error.

Let $X^r \subset \mathcal{H}$ be the span of the first r POD modes, and assume X^r is also contained in W. Let $P_r : \mathcal{H} \to \mathcal{H}$ be the orthogonal POD projection onto X^r, and let $\Pi_r^W : W \to W$ be the W-orthogonal projection onto X^r. Also, let s be the number of positive POD eigenvalues.

DEFINITION 4.6. *We say that the ROM discretization error Λ_r is*
- truly optimal *if there exists a constant C such that*

$$(4.26) \qquad \Lambda_r \leq C \Lambda_r^\star, \quad \Lambda_r^\star := \max_{1 \leq k \leq N} \|u^k - \Pi_r^W u^k\|_W^2,$$

- optimal-I *if there exists a constant C such that*

$$(4.27) \qquad \Lambda_r \leq C \Lambda_r^I, \quad \Lambda_r^I := \sum_{i=r+1}^{s} \lambda_i \|\varphi_i\|_W^2,$$

- optimal-II *if there exists a constant C such that*

$$(4.28) \qquad \Lambda_r \le C\Lambda_r^{II}, \quad \Lambda_r^{II} := \sum_{i=r+1}^{s} \lambda_i \|\varphi_i - \Pi_r^W \varphi_i\|_W^2.$$

The constant C above should be independent of all discretization parameters but may depend on the solution data and the problem data.

We note that the first two notions of optimality above are generalizations of definitions discussed in [28], while we believe the optimal-II definition is new. We discuss each type of optimality below.

Remark 4.7. Note that we do not consider the ROM discretization error for the initial condition, Λ_r^0, in these optimality definitions. These definitions can be modified to include the ROM initial condition error, if desired.

Truly optimal. Since Π_r^W is the W-orthogonal projection, the quantity $\Lambda_r^\star$ defined in (4.26) is the best possible pointwise POD data approximation error. As discussed in [28], this is the most natural definition of optimality; however, it may not be straightforward to evaluate the quantity $\Lambda_r^\star$ and compare it to the ROM discretization error bound Λ_r.

Optimal-I (optimal type I). Since it may not be easy to deal with the notion of truly optimal, Iliescu and Wang proposed the notion of Optimal-I in [28]. Optimal-I has the advantage of being simple to compute since Λ_r^I involves only the POD eigenvalues and modes. Optimal-I is also simple to interpret since from Lemma 2.4 we have

$$(4.29) \qquad \Lambda_r^I = \frac{1}{2N+1} \sum_{n=0}^{N} \|u^n - P_r u^n\|_W^2 + \frac{1}{2N+1} \sum_{n=1}^{N} \|\partial u^n - P_r \partial u^n\|_W^2 \,.$$

Therefore, Λ_r^I is the *total* POD projection error for all of the data using the POD projection P_r and the error norm W.

Optimal-II (optimal type II). The value of Λ_r^{II} is also relatively straightforward to compute, since it involves only POD eigenvalues, modes, and the projection Π_r^W. Also, by Lemma 2.4 we have

$$(4.30) \qquad \Lambda_r^{II} = \frac{1}{2N+1} \sum_{n=0}^{N} \left\|u^n - \Pi_r^W u^n\right\|_W^2 + \frac{1}{2N+1} \sum_{n=1}^{N} \left\|\partial u^n - \Pi_r^W \partial u^n\right\|_W^2 \,.$$

Since Π_r^W is the W-orthogonal projection, the quantity Λ_r^{II} is the best possible *total* POD data approximation error, and (4.29)–(4.30) imply

$$\Lambda_r^{II} \le \Lambda_r^I.$$

Optimal-II has the advantage of using a best possible POD approximation error, while also being relatively simple to compute and understand. Finally, we note that if $W = \mathcal{H}$, then $P_r = \Pi_r^W$ and therefore Optimal-I and Optimal-II are identical; however, Optimal-I and Optimal-II may be different if $\mathcal{H} \ne W$.

Comparing the optimality types. Since we are in the DQ case, the pointwise POD projection error result Theorem 3.7 implies that there exists a constant C such that

$$\Lambda_r^\star \le C\Lambda_r^{II}.$$

The above definitions, observations, and inequalities give the following result comparing the optimality types.

PROPOSITION 4.8. *The following hold:*
(i) *If the ROM discretization error is truly optimal, then it is Optimal-*II.
(ii) *If the ROM discretization error is Optimal-*II*, then it is Optimal-*I.
(iii) *If $\mathcal{H} = W$, then Optimal-*I *and Optimal-*II *are identical conditions.*
(iv) *If there exists a constant C such that*

$$(4.31) \qquad \|\varphi_i\|_W \leq C\|\varphi_i - \Pi_r^W \varphi_i\|_W, \quad r+1 \leq i \leq s,$$

*and if the ROM discretization error is Optimal-*I*, then it is Optimal-*II.

In general, we do not know if Optimal-II implies truly optimal; however, again, Λ_r^{II} is easier to deal with compared to $\Lambda_r^\star$. We also do not know in general if Optimal-I implies Optimal-II when $\mathcal{H} \neq W$. We discuss condition (4.31) below.

Remark 4.9 (the noDQ case). In the noDQ case, the same definitions of optimality can be used and Lemma 2.2 also gives interpretations of Λ_r^I and Λ_r^{II} as total POD projections errors in the W norm. As in the DQ case, Optimal-II implies Optimal-I, the two conditions are equivalent if $\mathcal{H} = W$, and Optimal-I with (4.31) implies Optimal-II.

However, as shown in Proposition 3.3, in general we cannot bound the pointwise POD projection error by a constant multiple of the total POD projection error, i.e., Assumption 3.1 is not always satisfied. Thus, we do not know if truly optimal implies Optimal-II. Furthermore, even if Assumption 3.1 is satisfied, the L^2 pointwise error estimate (4.20) in Remark 4.4 is not optimal in any sense, since the second term on its right-hand side contains $\|\nabla \varphi_i\|_{L^2}^2$ instead of $\|\varphi_i\|_{L^2}^2$.

Optimality of bounds in section 4.1. Next, we consider the optimality type of each pointwise in time error bound for the DQ case from section 4.1. Comparing the pointwise bounds in Lemmas 4.3 and 4.5 to the above optimality definitions gives the following result.

THEOREM 4.10. *For the pointwise error bounds in Lemma* 4.3 *with error norm $W = L^2$,*
(i) *if the L^2 POD basis is used (i.e., $\mathcal{H} = L^2$) and there exists a constant C such that*

$$(4.32) \qquad \|\varphi_i - R_r(\varphi_i)\|_{L^2} \leq C, \quad r+1 \leq i \leq s,$$

then the ROM discretization error in (4.11) *is Optimal-*I *(which is identical to Optimal-*II*);*
(ii) *if the H_0^1 POD basis is used (i.e., $\mathcal{H} = H_0^1$), then the ROM discretization error in* (4.13) *is Optimal-*I*;*
(iii) *if the H_0^1 POD basis is used (i.e., $\mathcal{H} = H_0^1$) and condition* (4.31) *is satisfied (with $W = L^2$), then the ROM discretization error in* (4.13) *is Optimal-*II.
For the pointwise error bounds in Lemma 4.5 *with error norm $W = H_0^1$,*
(iv) *if the L^2 POD basis is used (i.e., $\mathcal{H} = L^2$), then the ROM discretization error in* (4.21) *is Optimal-*II*;*
(v) *if the H_0^1 POD basis is used (i.e., $\mathcal{H} = H_0^1$), then the ROM discretization error in* (4.22) *is Optimal-*I *(which is identical to Optimal-*II*).*

Proof. Beginning with (i), the ROM discretization error from (4.11) is given by

$$(4.33) \qquad \Lambda_r = \sum_{i=r+1}^{s} \lambda_i^{DQ} \|\varphi_i - R_r(\varphi_i)\|_{L^2}^2.$$

By (4.32), the L^2 orthonormality of the POD basis, and the definition of Optimal-I it follows that

$$(4.34) \qquad \Lambda_r \leq C \sum_{i=r+1}^{s} \lambda_i^{DQ} = C \sum_{i=r+1}^{s} \lambda_i^{DQ} \|\varphi_i\|_{L^2}^2 = C\Lambda_r^I.$$

From Proposition 4.8 since $\mathcal{H} = W$ this is identical to Optimal-II.

For (ii) the ROM discretization error from (4.13) is given by

$$(4.35) \qquad \Lambda_r = \sum_{i=r+1}^{s} \lambda_i^{DQ} \|\varphi_i\|_{L^2}^2,$$

which is Optimal-I by definition.

Next, (iii) follows from (ii) and Proposition 4.8.

For (iv), the ROM discretization error in (4.21) is given by

$$(4.36) \qquad \Lambda_r = \sum_{i=r+1}^{s} \lambda_i^{DQ} \left(\|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \|\nabla(\varphi_i - R_r(\varphi_i))\|_{L^2}^2 \right).$$

Applying Poincaré's inequality to $\|\varphi_i - R_r(\varphi_i)\|_{L^2}^2$ shows that Λ_r is Optimal-II.

Finally, to prove (v) we use the fact that $P_r = R_r$ for $\mathcal{H} = H_0^1$, Poincaré's inequality, and the fact that $P_r \varphi_i = 0$ for $i > r$ to obtain

$$\Lambda_r = C \sum_{i=r+1}^{s} \lambda_i^{DQ} \left(\|\varphi_i - P_r(\varphi_i)\|_{L^2}^2 + \|\nabla(\varphi_i - P_r(\varphi_i))\|_{L^2}^2 \right) \leq C \sum_{i=r+1}^{s} \lambda_i^{DQ} \|\nabla\varphi_i\|_{L^2}^2,$$

which is Optimal-I by definition. Since $W = \mathcal{H} = H_0^1$, this is identical to Optimal-II by Proposition 4.8. □

The $W = L^2$ and $\mathcal{H} = H_0^1$ case suggests it may be possible for the ROM discretization error to be Optimal-I but not Optimal-II, since an additional assumption is required for Optimal-II. However, no other case shows a substantial difference between Optimal-I and Optimal-II. It is possible that further differences arise for other partial differential equations; we leave this to be investigated elsewhere.

We note that (4.31) and (4.32) are uniform boundedness type conditions for nonorthogonal POD projections. Indeed, for the case $W = \mathcal{H} = L^2$, the Ritz projection $R_r : L^2 \to L^2$ is not orthogonal (even though it is orthogonal when viewed as a mapping $R_r : H_0^1 \to H_0^1$). Thus, (4.32) is a uniform boundedness condition for a nonorthogonal POD projection. Furthermore, for the case $W = L^2$ and $\mathcal{H} = H_0^1$, we have $R_r\varphi_i = 0$ for $i > r$, and so (4.31) can be viewed as

$$(4.37) \qquad \|\varphi_i - R_r\varphi_i\|_{L^2} \leq C\|\varphi_i - \Pi_r^{L^2}\varphi_i\|_{L^2}, \quad r+1 \leq i \leq s.$$

Thus, (4.31) is a uniformly bounded comparison of a nonorthogonal POD projection with an orthogonal POD projection. These types of uniform boundedness conditions have been considered in [5, 28, 32, 39, 50, 56], but they are not well understood. We do not consider them further here; we leave them to be more fully explored elsewhere.

5. Numerical results. In this section, we investigate numerically Assumption 3.1. Specifically, we consider the following questions: (i) Is Assumption 3.1 satisfied? (ii) Is the pointwise in time projection error optimal? (iii) Is the pointwise in time ROM error optimal? To investigate these questions numerically, we use

the two counterexamples proposed in sections 3.1.1–3.1.2: counterexample 1, which was defined in (3.17), and counterexample 2, which was defined in (3.19). For each counterexample, we consider both the noDQ case (i.e., when the DQs are not used to construct the ROM basis; see section 2.1) and the DQ case (i.e., when the DQs are used to construct the ROM basis; see section 2.2).

Based on the theoretical results in sections 3.1 and 4.1, we expect the noDQ case to (i) violate Assumption 3.1 (see (3.4)); (ii) yield suboptimal pointwise projection errors (see (3.6) in Proposition 3.3); and (iii) yield suboptimal pointwise ROM errors (see (4.20)). In contrast, based on the theoretical results in sections 3.2 and 4.1, we expect the DQ case to (i) fulfill Assumption 3.1 (see Theorem 3.7); (ii) yield optimal pointwise projection errors (see Theorem 3.7); and (iii) yield optimal pointwise ROM errors (see 4.11).

In our numerical investigation, we use the one-dimensional heat equation (1.1), which was used in the theoretical development in section 4. For all the numerical experiments, we consider $\nu = 1$. We note that the time step, Δt, plays an important role in our theoretical and numerical investigation. Indeed, an $(N+1) = (T\Delta t^{-1}+1)$ factor determines the suboptimality of the pointwise projection and ROM error bounds for the noDQ case (see (3.6) and (4.20), respectively). Thus, in our numerical investigation it is desirable to consider as many Δt values as possible in order to study the asymptotic behavior of the error as Δt goes to zero. We note, however, that the two counterexamples that we investigate restrict the Δt values that we can consider. The reason is that while the two counterexamples yield ROM basis functions that are scaled versions of the snapshots (which is advantageous for the theoretical development), the treatment of their boundary conditions is somewhat delicate. Indeed, both counterexamples vanish at $x = 0$, but not at $x = 1$. To simplify the numerical treatment of the right boundary condition, we consider snapshots at Δt values for which $k\,\Delta t$ is an integer. This choice yields snapshots that vanish both at $x = 0$ and at $x = 1$, which allows for a straightforward ROM construction. To summarize, in our numerical investigation we strive to consider optimal k values that are large enough to ensure a large number of Δt values (while satisfying the restriction $k\,\Delta t \in \mathbb{N}$) and also low enough so that the numerical approximation is accurate.

Snapshot generation. Counterexamples 1 and 2 display a highly oscillatory behavior for the relatively large k values chosen (i.e., $k = 128$ and $k = 100$, respectively). Thus, to minimize the numerical error in generating the snapshots, we do not use a standard (e.g., FE) discretization. Instead, to construct the snapshots, we use the analytical forms of counterexamples 1 and 2 given in (3.17) and (3.19), respectively.

ROM construction. To construct the ROM basis, we collect equally spaced snapshots on the time interval $[0, 1]$ and $[0, 0.2]$ for counterexamples 1 and 2, respectively. Thus, the snapshot matrix K is $(N+1)$-dimensional in the noDQ case, and $(2N+1)$-dimensional in the DQ case, as explained in sections 2.1 and 2.2, respectively. To construct K, in (2.4) we use the standard Lagrange interpolant operator with respect to the FE nodes to interpolate the analytical solution of counterexamples 1 and 2. Next, we use K to build the ROM basis for the noDQ and DQ cases. We emphasize that, although K has different dimensions in the noDQ and DQ cases, to ensure a fair comparison, we use the same r value in all the numerical experiments. We construct the ROM operators by using the FE mass and stiffness matrices, which are obtained by using a linear FE spatial discretization with mesh size $\Delta h = 1/4096$. As the ROM initial condition, we use the L^2 projection of the initial condition in the noDQ case and the Ritz projection of the initial condition in the DQ case. We use these ROM operators to build the ROM, and run it over the time interval $[0, T]$ with the CN time discretization and the time step $\Delta t = T/N$.

5.1. Counterexample 1. In this section, we consider counterexample 1, which was proposed in (3.17) of section 3.1.1. In all the numerical experiments in this section, we consider $k = 128$ in (3.17). The numerical results are organized as follows. In section 5.1.1, for both the noDQ and the DQ case, we investigate numerically whether (i) Assumption 3.1 holds and (ii) the pointwise projection error is optimal. In section 5.1.2, for both the noDQ and the DQ case, we investigate numerically whether the pointwise ROM errors are optimal.

As explained in section 3.1.1, counterexample 1 was constructed to display the suboptimality of the pointwise projection and ROM bounds when $r = N$ and $t = t_N$. Thus, in our numerical investigation we also consider $r = N$ and $t = t_N$.

5.1.1. Pointwise projection error. In this section, we investigate numerically whether Assumption 3.1 holds. To this end, we monitor the magnitude of the projection error (1.11),

$$(5.1) \qquad \left\| \eta^{proj}(., t_n) \right\|_{L^2} = \left\| u(., t_n) - \sum_{i=1}^{N} \left(u(., t_n), \varphi_i \right)_{L^2} \varphi_i \right\|_{L^2}, \qquad n = 0, \ldots, N,$$

at all the time instances, and check whether there are large variations in its magnitude. Furthermore, for various Δt values, we investigate numerically whether the projection error (5.1) at the last time step is suboptimal (i.e., it has a suboptimal Δt^{-1} factor). Specifically, as shown in (3.3) for counterexample 1 in the noDQ case, the projection error at the last time step satisfies

$$(5.2) \qquad \left\| \eta^{proj}(., t_N) \right\|_{L^2}^2 = C_{proj}^{noDQ} \sum_{i=N+1}^{N+1} \lambda_i^{noDQ} \left\| \varphi_i \right\|_{L^2}^2,$$

where

$$(5.3) \qquad C_{proj}^{noDQ} = T \, \Delta t^{-1} + 1 = (N + 1).$$

Moreover, as shown in (3.24b) for counterexample 1 in the DQ case, the projection error at the last time step satisfies

$$(5.4) \qquad \left\| \eta^{proj}(., t_N) \right\|_{L^2}^2 \le C_{proj}^{DQ} \sum_{i=N+1}^{N+1} \lambda_i^{DQ} \left\| \varphi_i \right\|_{L^2}^2,$$

where

$$(5.5) \qquad C_{proj}^{DQ} = \mathcal{O}(1).$$

In this section, we investigate numerically the scalings (5.2) and (5.4).

noDQ case. In Table 1, for the noDQ case, we list the pointwise projection errors (5.1) at each time step. These results show that the pointwise projection error at the last time step is orders of magnitude higher than the pointwise projection error at the other time steps. Thus, we conclude that, in the noDQ case, counterexample 1 violates Assumption 3.1.

In Table 2, we list the scaling factor (5.2) for different Δt values. As expected from (5.3), these results show that the scaling factor is equal to $(N + 1)$. Thus, we conclude that, in the noDQ case, counterexample 1 yields suboptimal pointwise projection errors.

TABLE 1

Counterexample 1 (3.17), $\Delta t = 1/16$, *noDQ case: pointwise projection error* (5.1) *at each time step.*

n	$\lVert\eta^{proj}(.,t_n)\rVert_{L^2}$	n	$\lVert\eta^{proj}(.,t_n)\rVert_{L^2}$	n	$\lVert\eta^{proj}(.,t_n)\rVert_{L^2}$
0	$2.79e-08$	6	$2.11e-08$	12	$0.00e+00$
1	$2.24e-08$	7	$0.00e+00$	13	$1.49e-08$
2	$2.69e-08$	8	$1.67e-08$	14	$7.45e-09$
3	$7.45e-09$	9	$1.05e-08$	15	$1.67e-08$
4	$1.49e-08$	10	$2.11e-08$	16	$7.07e-01$
5	$1.83e-08$	11	$1.05e-08$		

TABLE 2

Counterexample 1 (3.17), *noDQ case: scaling factor* (5.2) *for different time step values.*

Δt	1/4	1/8	1/16	1/32	1/64	1/128
C_{proj}^{noDQ}	$5.0e+00$	$9.0e+00$	$1.7e+01$	$3.3e+01$	$6.5e+01$	$1.3e+02$

TABLE 3

Counterexample 1 (3.17), $\Delta t = 1/16$, *DQ case: pointwise projection error* (5.1) *at each time step.*

n	$\lVert\eta^{proj}(.,t_n)\rVert_{L^2}$	n	$\lVert\eta^{proj}(.,t_n)\rVert_{L^2}$	n	$\lVert\eta^{proj}(.,t_n)\rVert_{L^2}$
0	$1.7144e-01$	6	$1.7144e-01$	12	$1.7146e-01$
1	$1.7144e-01$	7	$1.7145e-01$	13	$1.7146e-01$
2	$1.7144e-01$	8	$1.7145e-01$	14	$1.7146e-01$
3	$1.7144e-01$	9	$1.7145e-01$	15	$1.7146e-01$
4	$1.7144e-01$	10	$1.7145e-01$	16	$1.7147e-01$
5	$1.7144e-01$	11	$1.7146e-01$		

TABLE 4

Counterexample 1 (3.17), *DQ case: scaling factor* (5.4) *for different time step values.*

Δt	1/4	1/8	1/16	1/32	1/64	1/128
$\mathcal{C}_{proj}^{DQ}$	$1.8e+00$	$1.9e+00$	$1.9e+00$	$2.0e+00$	$2.0e+00$	$2.0e+00$

DQ case. In Table 3, for the DQ case, we list the pointwise projection errors (5.1) at each time step. These results show that, in contrast with the noDQ case, the pointwise projection error at the last time step is of the same order of magnitude as the pointwise projection error at the other time steps. Thus, we conclude that, in the DQ case, counterexample 1 satisfies Assumption 3.1.

In Table 4, we list the scaling factor (5.4) for different time step values. As expected from (5.5), these results show that the scaling factor is bounded. Thus, we conclude that, in the DQ case, counterexample 1 yields optimal pointwise projection errors.

The numerical results in this section support the theoretical results in section 3. Specifically, counterexample 1 satisfies Assumption 3.1 in the DQ case but not in the noDQ case. Furthermore, the pointwise projection error at the last time step is optimal in the DQ case and suboptimal in the noDQ case.

5.1.2. Pointwise ROM error. In this section, we investigate whether the pointwise ROM error is suboptimal.

noDQ case. In the noDQ case, we investigate numerically the error estimate proved in (4.20):

(5.6)

$$\max_{1\leq k\leq N}\|e^k\|_{L^2}^2 = \mathcal{O}\left((N+1)\sum_{i=N+1}^{N+1}\lambda_i^{noDQ}\|\varphi_i\|_{L^2}^2 + \Delta t^4 + \sum_{i=N+1}^{N+1}\lambda_i^{noDQ}\|\nabla\varphi_i\|_{L^2}^2\right).$$

We note that, since the ROM initial condition is the L^2 projection of the initial condition, the term $\|\phi_r^0\|_{L^2}^2$ in (4.20) vanishes in (5.6). As explained in Remark 4.4, the error bound (5.6) is suboptimal with respect to the time step due to the factor $(N+1) = (\Delta t^{-1} + 1)$ in the first term on the right-hand side. To investigate numerically the suboptimality of the error bound (5.6), in Table 5 we list the ratio

$$(5.7) \qquad C_{rom}^{noDQ} = \left(\max_{1\leq k\leq N}\|e^k\|_{L^2}^2\right) / \left((N+1)\sum_{i=N+1}^{N+1}\lambda_i^{noDQ}\|\varphi_i\|_{L^2}^2 \right.$$
$$\left. + \Delta t^4 + \sum_{i=N+1}^{N+1}\lambda_i^{noDQ}\|\nabla\varphi_i\|_{L^2}^2\right).$$

The results in Table 5 show that the ratio (5.7) is bounded from below. Thus, we conclude that the pointwise ROM error in the noDQ case is suboptimal.

To investigate the sensitivity of our numerical results with respect to k (i.e., the level of oscillations in counterexample 1), in Table 6 we list the ratio (5.7) for $k = 8$. The results in Table 6 confirm the results in Table 5, i.e., the pointwise ROM error in the noDQ case is suboptimal.

DQ case. In the DQ case, we investigate numerically the error estimate proved in (4.11):

$$(5.8) \qquad \max_{1\leq k\leq N}\|e^k\|_{L^2}^2 = \mathcal{O}\left(\sum_{i=N+1}^{N+1}\lambda_i^{DQ}\|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \Delta t^4\right).$$

We note that the error bound (5.8) is optimal. In Table 7, we list the ratio

$$(5.9) \qquad C_{rom}^{DQ} = \left(\max_{1\leq k\leq N}\|e^k\|_{L^2}^2\right) / \left(\sum_{i=N+1}^{N+1}\lambda_i^{DQ}\|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \Delta t^4\right).$$

The results in Table 7 show that the ratio (5.9), while increasing, seems to be bounded, as predicted by (5.8).

The increase of C_{rom}^{DQ} in Table 7 is due to the highly oscillatory character of counterexample 1 in (3.17), which makes the ROM simulation in the DQ case challenging.

TABLE 5
Counterexample 1 (3.17), noDQ case: ratio (5.7) for different time step values.

Δt	1/4	1/8	1/16	1/32	1/64	1/128
$\mathcal{C}_{rom}^{noDQ}$	$3.0e-04$	$1.8e-04$	$1.0e-04$	$2.0e-04$	$7.6e-04$	$7.9e-04$

TABLE 6
Counterexample 1 (3.17), $k = 8$, noDQ case: ratio (5.7) for different time step values.

Δt	1/2	1/4	1/8
$\mathcal{C}_{rom}^{noDQ}$	$3.75e-03$	$6.371e-03$	$1.13-02$

TABLE 7
Counterexample 1 (3.17), *DQ case: ratio* (5.9) *for different time step values.*

Δt	1/4	1/8	1/16	1/32	1/64	1/128
C_{rom}^{DQ}	$7.8e-02$	$1.3e-01$	$2.0e-01$	$3.5e-01$	$5.3e-01$	$8.7e-01$

TABLE 8
Counterexample 1 (3.17), $k = 8$, *and DQ case: ratio* (5.9) *for different time step values.*

Δt	1/2	1/4	1/8
$\mathcal{C}_{rom}^{DQ}$	$4.73e-01$	$5.92e-01$	$2.55e-01$

To alleviate the highly oscillatory behavior of counterexample 1, we keep all the parameters unchanged and choose a lower k value (i.e., $k = 8$) in (3.17), which yields a solution with fewer oscillations. In Table 8, we list the ratio (5.9) for $k = 8$. The results in Table 8 show that the ratio (5.9) is bounded, as predicted by (5.8).

The numerical results in this section support the theoretical results in section 4. Specifically, for counterexample 1, the pointwise ROM error is optimal in the DQ case, and suboptimal in the noDQ case.

5.2. Counterexample 2. In this section, we consider counterexample 2, which was proposed in (3.19) of section 3.1.2. In all the numerical experiments in this section, we consider $k = 100$, $\delta = 0.01$, and $\alpha = 1$ in (3.19). The numerical results are organized as follows. In section 5.2.1, for both the noDQ and the DQ case, we investigate numerically whether (i) Assumption 3.1 holds; and (ii) the pointwise projection error is optimal. In section 5.2.2, for both the noDQ and the DQ case, we investigate numerically whether the pointwise ROM errors are optimal.

As explained in section 3.1.2, counterexample 2 was constructed to display the suboptimality of the pointwise projection and ROM error bounds for any r values. In our numerical investigation, we consider general r and $t = t_k$ values for both the pointwise projection error and the pointwise ROM error.

5.2.1. Pointwise projection error. In this section, we investigate numerically whether Assumption 3.1 holds. To this end, for various Δt values, we investigate numerically whether the projection error (5.10) at various time instances is suboptimal.

$$(5.10) \qquad \left\| \eta^{proj}(.,t_r) \right\|_{L^2} = \left\| u(.,t_r) - \sum_{i=1}^{r} \left(u(.,t_r), \varphi_i \right)_{L^2} \varphi_i \right\|_{L^2}, \qquad r = 1, \ldots, N,$$

Specifically, as shown in (3.8) in Proposition 3.3 for counterexample 2 in the noDQ case, for fixed r values, the projection error at $t = t_r$ satisfies

$$(5.11) \qquad \left\| \eta^{proj}(.,t_r) \right\|_{L^2}^2 = C_{proj}^{noDQ} (N+1) \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} \left\| \varphi_i \right\|_{L^2}^2,$$

where

$$(5.12) \qquad C_{proj}^{noDQ} \geq \frac{\min\{1, \gamma\}}{2}.$$

Moreover, as shown in (3.24b) for counterexample 2 in the DQ case, the projection error at various time instances satisfies

TABLE 9

Counterexample 2 (3.19), $r = 4$, and noDQ case: scaling factor (5.11) for different time step values.

Δt	0.05	0.04	0.02	0.01
$\mathcal{C}_{proj}^{noDQ}$	$1.00e + 00$	$9.82e - 01$	$8.65e - 01$	$6.32e - 01$

TABLE 10

Counterexample 2 (3.19), $r = 4$, and DQ case: scaling factor (5.13) for different time step values.

Δt	0.05	0.04	0.02	0.01
$\mathcal{C}_{proj}^{DQ}$	$1.83e + 00$	$1.76e - 02$	$8.32e - 03$	$3.84e - 03$

$$(5.13) \qquad \max_{0 \leq k \leq N} \left\| u(.,t_k) - \sum_{i=1}^{r} \left(u(.,t_k), \varphi_i \right)_{L^2} \varphi_i \right\|_{L^2}^2 \leq C_{proj}^{DQ} \sum_{i=r+1}^{d} \lambda_i^{\mathrm{DQ}} \|\varphi_i\|_{L^2},$$

where C_{proj}^{DQ} is bounded from above. In this section, we investigate numerically the scalings (5.11)–(5.12) and (5.13).

noDQ case. In Table 9, for $r = 4$, we list the scaling factor $\mathcal{C}_{proj}^{DQ}$ in (5.11) for different time step values. As expected from (5.12), these results show that the scaling factor is bounded from below. Thus, we conclude that, in the noDQ case, counterexample 2 yields suboptimal pointwise projection errors.

DQ case. In Table 10, for $r = 4$, we list the scaling factor (5.13) for different time step values. As expected, these results show that the scaling factor is bounded from above. Thus, we conclude that, in the DQ case, counterexample 2 yields optimal pointwise projection errors.

The numerical results in this section support the theoretical results in section 3. Specifically, for a generic r value, counterexample 2 satisfies Assumption 3.1 in the DQ case but not in the noDQ case. Furthermore, the pointwise projection error is optimal in the DQ case and suboptimal in the noDQ case.

5.2.2. Pointwise ROM error. In this section, we investigate whether the pointwise ROM error is suboptimal. We note that the time evolution of the analytical solution in counterexample 2 (which is displayed in Figure 1) prompted us to make the following parameter choices in the numerical investigation of the pointwise ROM error. Since the magnitude of the analytical solution is significant on the time interval $[0, 0.04]$ and almost negligible on the time interval $[0.04, 0.2]$, we decided to compute the pointwise ROM errors for both the noDQ and the DQ case on the time interval $[0, 0.05]$. Furthermore, since the DQ ROM basis functions with large indices are very oscillatory, we decided to use low r values in order to avoid numerical instabilities.

noDQ case. In the noDQ case, we investigate numerically the error estimate proved in (4.20):

$$(5.14)$$

$$\max_{1 \leq k \leq N} \|e^k\|_{L^2}^2 = \mathcal{O} \left((N + 1) \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} \|\varphi_i\|_{L^2}^2 + \Delta t^4 + \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} \|\nabla \varphi_i\|_{L^2}^2 \right).$$

We note that since the ROM initial condition is the L^2 projection of the initial condition, the term $\|\phi_r^0\|_{L^2}^2$ in (4.20) vanishes in (5.14). As explained in Remark 4.4,
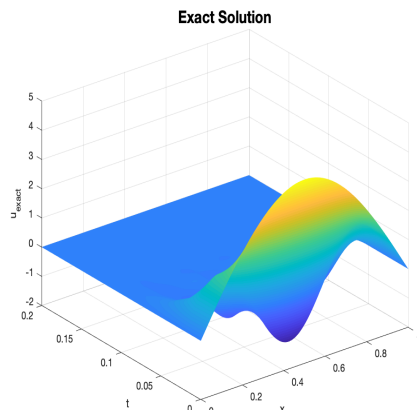
Exact Solution

FIG. 1. *Counterexample* 2 (3.19), *FOM plot:* $h = 1/4096$ *and* $\Delta t = 0.02$.

TABLE 11
Counterexample 2 (3.19) *and noDQ case: ratio* (5.15) *for fixed time step* $\Delta t = 0.01$ *and different* r *values.*

r	1	2	3	4	5	6
$\mathcal{C}_{rom}^{noDQ}$	$1.7e-01$	$9.8e-02$	$1.1e-01$	$2.2e-01$	$4.4e-01$	$9.2e-01$

the error bound (5.14) is suboptimal with respect to the time step due to the factor $(N + 1) = (T\Delta t^{-1} + 1)$ in the first term on the right-hand side. To investigate numerically the suboptimality of the error bound (5.14), in Table 11 we list the ratio (5.15) for fixed Δt values and various r values. The ratios in Table 11 are bounded from below. Thus, we conclude that the pointwise ROM error in the noDQ case is suboptimal.

$$(5.15) \qquad C_{rom}^{noDQ} = \left(\max_{1 \le k \le N} \|e^k\|_{L^2}^2 \right) / \left((N+1) \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} \|\varphi_i\|_{L^2}^2 \right.$$
$$\left. + \Delta t^4 + \sum_{i=r+1}^{N+1} \lambda_i^{noDQ} \|\nabla \varphi_i\|_{L^2}^2 \right).$$

DQ case. In the DQ case, we investigate numerically the error estimate proved in (4.11):

$$(5.16) \qquad \max_{1 \le k \le N} \|e^k\|_{L^2}^2 = \mathcal{O}\left(\sum_{i=r+1}^{N+1} \lambda_i^{DQ} \|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \Delta t^4 \right).$$

To investigate numerically the optimality of the error bound (5.16), in Table 12 we list the ratio (5.17) for fixed Δt values and various r values. The ratios in Table 12 are bounded. Thus, we conclude that the pointwise ROM error in the DQ case is optimal.

$$(5.17) \qquad C_{rom}^{DQ} = \left(\max_{1 \le k \le N} \|e^k\|_{L^2}^2 \right) / \left(\sum_{i=r+1}^{N+1} \lambda_i^{DQ} \|\varphi_i - R_r(\varphi_i)\|_{L^2}^2 + \Delta t^4 \right).$$

TABLE 12
Counterexample 2 (3.19) and DQ case: ratio (5.17) for fixed time step $\Delta t = 0.01$ and different r values.

r	1	2	3	4	5	6
$\mathcal{C}_{rom}^{DQ}$	$2.9e - 03$	$4.0e - 03$	$4.9e - 03$	$5.7e - 03$	$1.0e - 02$	$2.9e - 02$

The numerical results in this section support the theoretical results in section 4. Specifically, for counterexample 2, the pointwise ROM error is optimal in the DQ case and suboptimal in the noDQ case.

6. Conclusions. In this paper, we resolved several theoretical issues dealing with the optimality of pointwise in time error bounds for POD-ROMs of the heat equation. In particular, we studied the role played by the DQs in the optimality of pointwise POD error bounds with respect to (i) the time discretization error and (ii) the ROM discretization error.

First, in the noDQ case (i.e., when the DQs are not used to construct the POD basis), we proved that the error bound is suboptimal not only with respect to the ROM discretization (as shown in [28]) but also with respect to the time discretization. Specifically, in Proposition 3.3 we constructed two classes of analytical examples, and we proved that these examples violate Assumption 3.1 and yield suboptimal (with respect to the time discretization) pointwise projection error bounds. Furthermore, we noted that these suboptimal pointwise projection error bounds yield suboptimal ROM error bounds (see Remark 4.4). Finally, we illustrated the suboptimality of the pointwise projection and ROM error bounds in the numerical simulation of the heat equation.

Our second main contribution is Theorem 3.7, where we proved that, in the DQ case (i.e., when the DQs are used to construct the POD basis), Assumption 3.1 is always satisfied. To prove Theorem 3.7, in Lemma 3.6 we first proved a discrete time Sobolev inequality for the DQ case. Next, in section 4, we used Theorem 3.7 to prove pointwise ROM error bounds that are optimal with respect to both the ROM discretization error and the time discretization error in the DQ case. In section 5, we illustrated the optimality of the pointwise projection and error bounds in the numerical simulation of the heat equation.

Our third main contribution is that, in Definition 4.6, we proposed a new definition for the optimality of pointwise in time ROM discretization errors. In section 4.2, we carefully discussed the relationship between this new optimality definition and the other two optimality definitions in current use. In Theorem 4.10, for two of the three optimality definitions, we showed that the DQ case yields optimal bounds, whereas the noDQ case yields suboptimal error bounds.

Our theoretical and numerical investigations (see also [28, 34, 50]) show that the DQs are needed to prove optimal pointwise in time error bounds. There are, however, several research directions that need to be investigated.

At a theoretical level, the uniform boundedness type conditions for nonorthogonal POD projections considered in Proposition 4.8 and Theorem 4.10 are important in proving some of the optimal pointwise ROM error bounds. These type of uniform boundedness conditions have been studied both theoretically and numerically in [5, 28, 32, 39, 50, 56], but they are not well understood. Further investigation of these conditions is needed. Additionally, at a theoretical level we considered optimal uniform estimates only for the heat equation. How these estimates will extend to more complicated nonlinear PDEs (e.g., the Navier–Stokes equations) is an open problem.

In this paper, we considered equally spaced snapshots to construct the POD basis. The POD adaptivity in time (see, e.g., [1, 26, 36, 42] and the survey in [16]) aims at choosing snapshot time instances that are optimal in some sense (e.g., such that the error between the ROM and FOM trajectories is minimized [36]). The effect of POD adaptivity in time on the optimality of error bounds in the noDQ and DQ cases should also be investigated.

At a numerical level, further investigation of the role of DQs in practical computations is needed. In this paper, we focused exclusively on the optimality of the rates of convergence of ROM error bounds. We emphasize, however, that we did not address the size of the ROM error. In our numerical investigation, the size of the ROM error was sometimes lower in the noDQ case and other times lower in the DQ case. Overall, the size of the ROM error was of the same order in the noDQ and DQ cases (results are not included). We note that the current literature yields similar qualitative conclusions: In some references [24, 33], the ROM error is lower in the DQ case than in the noDQ case; in other references [28, 32, 34], the situation is reversed. Further investigation of the size of the ROM error in the noDQ and DQ cases is needed.

Acknowledgment. We thank the three anonymous reviewers for their constructive comments and suggestions, which improved the manuscript.

REFERENCES

[1] A. ALLA, C. GRÄSSLE, AND M. HINZE, *A posteriori snapshot location for POD in optimal control of linear parabolic equations*, ESAIM Math. Model. Numer. Anal., 52 (2018), pp. 1847–1873.

[2] M. AZAÏEZ, T. CHACÓN REBOLLO, AND S. RUBINO, *A Streamline Derivative Projection-Based POD-ROM for Convection-Dominated Flows. Part* I: *Numerical Analysis*, https://arxiv.org/abs/1711.09780v1, 2017.

[3] M. BARRAULT, Y. MADAY, N. C. NGUYEN, AND A. T. PATERA, *An 'empirical interpolation' method: Application to efficient reduced-basis discretization of partial differential equations*, C. R. Acad. Sci. Paris Ser. I, 339 (2004), pp. 667–672.

[4] K. CARLBERG AND C. FARHAT, *A low-cost, goal-oriented compact proper orthogonal decomposition basis for model reduction of static systems*, Internat. J. Numer. Methods Engrg., 86 (2011), pp. 381–402.

[5] D. CHAPELLE, A. GARIAH, AND J. SAINTE-MARIE, *Galerkin approximation with proper orthogonal decomposition: New error estimates and illustrative examples*, ESAIM Math. Model. Numer. Anal., 46 (2012), pp. 731–757.

[6] S. CHATURANTABUT AND D. C. SORENSEN, *A state space error estimate for POD-DEIM nonlinear model reduction*, SIAM J. Numer. Anal., 50 (2012), pp. 46–63.

[7] Y. CHOI, D. COOMBS, AND R. ANDERSON, *SNS: A solution-based nonlinear subspace method for time-dependent model order reduction*, SIAM J. Sci. Comput., 42 (2020), pp. A1116–A1146.

[8] V. DECARIA, T. ILIESCU, W. LAYTON, M. MCLAUGHLIN, AND M. SCHNEIER, *An artificial compression reduced order model*, SIAM J. Numer. Anal., 58 (2020), pp. 565–589.

[9] F. G. EROGLU, S. KAYA, AND L. G. REBHOLZ, *A modular regularized variational multiscale proper orthogonal decomposition for incompressible flows*, Comput. Methods Appl. Mech. Engrg., 325 (2017), pp. 350–368.

[10] F. G. EROGLU, S. KAYA, AND L. G. REBHOLZ, *POD-ROM for the Darcy-Brinkman equations with double-diffusive convection*, J. Numer. Math., 27 (2019), pp. 123–139, https://doi.org/10.1515/jnma-2017-0122.

[11] L. C. EVANS, *Partial Differential Equations*, 2nd ed., Grad. Stud. Math. 19, AMS, Providence, RI, 2010, https://doi.org/10.1090/gsm/019.

[12] L. FICK, Y. MADAY, A. T. PATERA, AND T. TADDEI, *A stabilized POD model for turbulent flows over a range of Reynolds numbers: Optimal parameter sampling and constrained projection*, J. Comput. Phys., 371 (2018), pp. 214–243.

[13] P. Galán del Sastre and R. Bermejo, *Error estimates of proper orthogonal decomposition eigenvectors and Galerkin projection for a general dynamical system arising in fluid models*, Numer. Math., 110 (2008), pp. 49–81.

[14] S. Giere, T. Iliescu, V. John, and D. Wells, *SUPG reduced order models for convection-dominated convection-diffusion-reaction equations*, Comput. Methods Appl. Mech. Engrg., 289 (2015), pp. 454–474.

[15] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*, Springer Ser. Comput. Math. 5, Springer, Berlin, 1986, https://doi.org/10.1007/978-3-642-61623-5.

[16] C. Grässle, *Adaptivity in Model Order Reduction with Proper Orthogonal Decomposition*, Ph.D. thesis, University of Hamburg, 2019.

[17] M. A. Grepl and A. T. Patera, *A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations*, ESAIM Math. Model. Numer. Anal., 39 (2005), pp. 157–181.

[18] M. Gubisch and S. Volkwein, *Proper orthogonal decomposition for linear-quadratic optimal control*, in Model Reduction and Approximation, Comput. Sci. Eng. 15, SIAM, Philadelphia, 2017, pp. 3–63, https://doi.org/10.1137/1.9781611974829.ch1.

[19] M. Gunzburger, N. Jiang, and M. Schneier, *An ensemble-proper orthogonal decomposition method for the nonstationary Navier-Stokes equations*, SIAM J. Numer. Anal., 55 (2017), pp. 286–304.

[20] S. Herkt, M. Hinze, and R. Pinnau, *Convergence analysis of Galerkin POD for linear second order evolution equations*, Electron. Trans. Numer. Anal., 40 (2013), pp. 321–337.

[21] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, Springer, Berlin, 2015.

[22] S. Hijazi, G. Stabile, A. Mola, and G. Rozza, *Data-driven POD-Galerkin reduced order model for turbulent flows*, J. Comput. Phys., (2020), 109513.

[23] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, UK, 1996.

[24] D. Hömberg and S. Volkwein, *Control of laser surface hardening by a reduced-order approach using proper orthogonal decomposition*, Math. Comput. Model., 38 (2003), pp. 1003–1028.

[25] C. Homescu, L. R. Petzold, and R. Serban, *Error estimation for reduced-order models of dynamical systems*, SIAM J. Numer. Anal., 43 (2005), pp. 1693–1714.

[26] R. H. W. Hoppe and Z. Liu, *Snapshot location by error equilibration in proper orthogonal decomposition for linear and semilinear parabolic partial differential equations*, J. Numer. Math., 22 (2014), pp. 1–32.

[27] T. Iliescu and Z. Wang, *Variational multiscale proper orthogonal decomposition: Convection-dominated convection-diffusion-reaction equations*, Math. Comp., 82 (2013), pp. 1357–1378.

[28] T. Iliescu and Z. Wang, *Are the snapshot difference quotients needed in the proper orthogonal decomposition?*, SIAM J. Sci. Comput., 36 (2014), pp. A1221–A1250.

[29] T. Iliescu and Z. Wang, *Variational multiscale proper orthogonal decomposition: Navier-Stokes equations*, Numer. Methods Partial Differential Equations, 30 (2014), pp. 641–663.

[30] K. Ito and S. S. Ravindran, *A reduced-order method for simulation and control of fluid flows*, J. Comput. Phys., 143 (1998), pp. 403–425.

[31] B. Jin and Z. Zhou, *An analysis of Galerkin proper orthogonal decomposition for subdiffusion*, ESAIM Math. Model. Numer. Anal., 51 (2017), pp. 89–113, https://doi.org/10.1051/m2an/2016017.

[32] K. Kean and M. Schneier, *Error analysis of supremizer pressure recovery for POD based reduced-order models of the time-dependent Navier–Stokes equations*, SIAM J. Numer. Anal., 58 (2020), pp. 2235–2264, https://doi.org/10.1137/19M128702X.

[33] T. Kostova-Vassilevska and G. M. Oxberry, *Model reduction of dynamical systems by proper orthogonal decomposition: Error bounds and comparison of methods using snapshots from the solution and the time derivatives*, J. Comput. Appl. Math., 330 (2018), pp. 553–573.

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

[35] K. Kunisch and S. Volkwein, *Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics*, SIAM J. Numer. Anal., 40 (2002), pp. 492–515 (electronic).

[36] K. Kunisch and S. Volkwein, *Optimal snapshot location for computing POD basis functions*, ESAIM Math. Model. Numer. Anal., 44 (2010), pp. 509–529.

[37] W. J. Layton, *Introduction to the Numerical Analysis of Incompressible Viscous Flows*, Comput. Sci. Eng. 6, SIAM, Philadelphia, 2008.

[38] F. Leibfritz and S. Volkwein, *Numerical feedback controller design for PDE systems using model reduction: Techniques and case studies*, in Real-Time PDE-Constrained Optimization, Comput. Sci. Eng. 3, SIAM, Philadelphia, 2007, pp. 53–72, https://doi.org/10.1137/1.9780898718935.ch3.

[39] S. Locke and J. Singler, *New proper orthogonal decomposition approximation theory for PDE solution data*, SIAM J. Numer. Anal., 58 (2020), pp. 3251–3285, https://doi.org/10.1137/19M1297002.

[40] Z. Luo, J. Chen, I. M. Navon, and X. Yang, *Mixed finite element formulation and error estimates based on proper orthogonal decomposition for the nonstationary Navier–Stokes equations*, SIAM J. Numer. Anal., 47 (2008), pp. 1–19.

[41] M. Mohebujjaman, L. G. Rebholz, X. Xie, and T. Iliescu, *Energy balance and mass conservation in reduced order models of fluid flows*, J. Comput. Phys., 346 (2017), pp. 262–277.

[42] G. M. Oxberry, T. Kostova-Vassilevska, W. Arrighi, and K. Chand, *Limited-memory adaptive snapshot selection for proper orthogonal decomposition*, Internat. J. Numer. Methods Engrg., 109 (2017), pp. 198–217.

[43] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced Basis Methods for Partial Differential Equations: An Introduction*, Unitext 92, Springer, Berlin, 2015.

[44] M. Rathinam and L. R. Petzold, *A new look at proper orthogonal decomposition*, SIAM J. Numer. Anal., 41 (2003), pp. 1893–1925.

[45] 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*, Arch. Comput. Methods Eng., 15 (2008), pp. 229–275.

[46] S. Rubino, *A streamline derivative POD-ROM for advection-diffusion-reaction equations*, ESAIM Proc., 64 (2018), pp. 121–136.

[47] E. Sachs and M. Schu, *A priori error estimates for reduced order models in finance*, ESAIM: Math. Model. Numer. Anal., 47 (2013), pp. 449–469.

[48] J. Shen, J. R. Singler, and Y. Zhang, *HDG-POD reduced order model of the heat equation*, J. Comput. Appl. Math., 362 (2019), pp. 663–679, https://doi.org/10.1016/j.cam.2018.09.031.

[49] J. R. Singler, *Convergent snapshot algorithms for infinite-dimensional Lyapunov equations*, IMA J. Numer. Anal., 31 (2011), pp. 1468–1496, https://doi.org/10.1093/imanum/drq028.

[50] J. R. Singler, *New POD error expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs*, SIAM J. Numer. Anal., 52 (2014), pp. 852–876.

[51] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 2006.

[52] K. Urban and A. T. Patera, *A new error bound for reduced basis approximation of parabolic partial differential equations*, C. R. Acad. Sci. Paris Sér. I Math., 350 (2012), pp. 203–207.

[53] K. Veroy and A. T. Patera, *Certified real-time solution of the parametrized steady incompressible Navier–Stokes equations: Rigorous reduced-basis a posteriori error bounds*, Internat. J. Numer. Methods Fluids, 47 (2005), pp. 773–788.

[54] S. Volkwein, *Proper Orthogonal Decomposition: Theory and Reduced-order Modelling*, Lecture Notes, University of Konstanz, 2013, http://www.math.uni-konstanz.de/numerik/personen/volkwein/teaching/POD-Book.pdf.

[55] M. F. Wheeler, *A priori L_2 error estimates for Galerkin approximations to parabolic partial differential equations*, SIAM J. Numer. Anal., 10 (1973), pp. 723–759.

[56] X. Xie, D. Wells, Z. Wang, and T. Iliescu, *Numerical analysis of the Leray reduced order model*, J. Comput. Appl. Math., 328 (2018), pp. 12–29.

[57] M. Yano, *Discontinuous Galerkin reduced basis empirical quadrature procedure for model reduction of parametrized nonlinear conservation laws*, Adv. Comput. Math., 45 (2019), pp. 2287–2320.

[58] C. Zerfas, L. G. Rebholz, M. Schneier, and T. Iliescu, *Continuous data assimilation reduced order models of fluid flow*, Comput. Methods Appl. Mech. Engrg., 357 (2019), 112596.

[59] S. Zhu, L. Dedè, and A. Quarteroni, *Isogeometric analysis and proper orthogonal decomposition for parabolic problems*, Numer. Math., 135 (2017), pp. 333–370, https://doi.org/10.1007/s00211-016-0802-5.

[60] R. Zimmermann, *Gradient-enhanced surrogate modeling based on proper orthogonal decomposition*, J. Comput. Appl. Math., 237 (2013), pp. 403–418.