SIAM J. SCI. COMPUT.
Vol. 42, No. 5, pp. A2837–A2864

STABILITY OF DISCRETE EMPIRICAL INTERPOLATION AND GAPPY PROPER ORTHOGONAL DECOMPOSITION WITH RANDOMIZED AND DETERMINISTIC SAMPLING POINTS*

BENJAMIN PEHERSTORFER†, ZLATKO DRMAČ‡, AND SERKAN GUGERCIN§

Abstract. This work investigates the stability of (discrete) empirical interpolation for nonlinear model reduction and state field approximation from measurements. Empirical interpolation derives approximations from a few samples (measurements) via interpolation in low-dimensional spaces. It has been observed that empirical interpolation can become unstable if the samples are perturbed due to, e.g., noise, turbulence, and numerical inaccuracies. The main contribution of this work is a probabilistic analysis that shows that stable approximations are obtained if samples are randomized and if more samples than dimensions of the low-dimensional spaces are used. Oversampling, i.e., taking more sampling points than dimensions of the low-dimensional spaces, leads to approximations via regression and is known under the name of gappy proper orthogonal decomposition. Building on the insights of the probabilistic analysis, a deterministic sampling strategy is presented that aims to achieve lower approximation errors with fewer points than randomized sampling by taking information about the low-dimensional spaces into account. Numerical results of reconstructing velocity fields from noisy measurements of combustion processes and model reduction in the presence of noise demonstrate the instability of empirical interpolation and the stability of gappy proper orthogonal decomposition with oversampling.

Key words. model reduction, empirical interpolation, gappy proper orthogonal decomposition, noisy observations, nonlinear model reduction, randomized model reduction

AMS subject classifications. 65M22, 65N22, 65F25, 65F30

DOI. 10.1137/19M1307391

1. Introduction.

Model reduction seeks to construct reduced systems that provide accurate approximations of the solutions of large-scale systems of equations with significantly reduced computational cost [7]. In projection-based model reduction, the reduced systems are obtained via (Petrov–)Galerkin projection of the full-system equations onto low-dimensional—reduced—subspaces of the high-dimensional solution spaces corresponding to the full systems. If the large-scale systems contain nonlinear equations, then projection of the full-system equations onto reduced spaces typically is insufficient to obtain reduced systems that are computationally cheaper to solve than the full systems, because the nonlinear terms entail computations with costs that scale with the number of the degrees of freedom of the full system. The empirical interpolation method (EIM) [6, 27, 28] and its discrete counterpart, the discrete empirical interpolation method (DEIM) [14, 17], provide one solution to this

*Submitted to the journal's Methods and Algorithms for Scientific Computing section December 17, 2019; accepted for publication (in revised form) July 13, 2020; published electronically September 21, 2020.

https://doi.org/10.1137/19M1307391

Funding: The work of the first author was partially supported by the Air Force Center of Excellence on Multi-Fidelity Modeling of Rocket Combustor Dynamics through grant FA9550-17-1-0195. The work of the second author was partially supported by the Croatian Science Foundation through grant IP-2019-04-6268. The work of the third author was partially supported by the NSF through grants DMS-1522616 and DMS-1819110.

†Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (pehersto@cims.nyu.edu).

‡Faculty of Science, Department of Mathematics, University of Zagreb, 10000 Zagreb, Croatia (drmac@math.hr).

§Department of Mathematics and Computational Modeling and Data Analytics Division, Academy of Integrated Science, Virginia Tech, Blacksburg, VA 24061-0123 (gugercin@vt.edu).

problem by approximating the nonlinear terms of the nonlinear equations via sparse sampling. The nonlinear terms are evaluated at a few interpolation points—sampling points—and then all other components of the nonlinear terms are approximated via interpolation in low-dimensional subspaces. However, approximations via (D)EIM have been shown to suffer from instabilities in certain situations; see, e.g., [2, 21, 43]. Localization [19, 33] and adaptation [32, 34] of the low-dimensional subspaces have been proposed as possible remedies. Another remedy that has been reported in the literature, and that typically is easier to implement in practice than localization and adaptation, is "oversampling" empirical interpolation so that the nonlinear terms are approximated via regression rather than via interpolation, which goes under the name of gappy proper orthogonal decomposition (GappyPOD) in the model reduction literature [2, 4, 11, 43, 45]. In this work, we consider the specific case where only noisy samples—observations—of the nonlinear terms are available and where (D)EIM has been shown to be unstable; see, e.g., [2]. We provide a probabilistic analysis that shows that GappyPOD with randomized samples leads to stable approximations in the presence of noise if more sampling points than basis vectors are used.

Approximations based on regression, rather than interpolation, have been investigated in the context of model reduction. Missing point estimation (MPE) [3, 4] relies on GappyPOD [20] to approximate nonlinear terms in model reduction. Several sampling point selection algorithms have been proposed for MPE and GappyPOD. The work [42] formulates point selection as a sensor placement problem and proposes a greedy approach to find an approximate solution. Detailed analyses of point selection for MPE, and screening approaches to speed up point selection, are provided in [4]. The work by Zimmermann and Willcox [45] introduces a sampling strategy for MPE that is based on approximating eigenvalues for selecting sampling points and demonstrates that oversampling achieves higher accuracies in numerical experiments in computational fluid dynamics than MPE without oversampling. We will arrive at a special case of the approach presented in [45] via perturbation bounds on eigenvalues introduced in [26]. Carlberg et al. [12, 13] introduce the Gauss–Newton with approximated tensors (GNAT) method that is based on Petrov–Galerkin projection and approximates the nonlinear terms via low-cost least-squares problems as in GappyPOD. The GNAT method and its performance based on regression has been investigated in the thesis [11], where a greedy-based deterministic sampling strategy for selecting sampling points has been proposed. Zhou [43] introduces a deterministic sampling strategy for GappyPOD that exploits the dependency of the degrees of freedom of the full system to select sampling points. Regression via GappyPOD is then applied to multi-scale problems, where Zhou's sampling strategy with GappyPOD achieves lower errors than DEIM via interpolation. The adaptive DEIM, which adapts the DEIM space from sparse samples of the nonlinear terms, is based on regression [32, 34, 44], even though regression is used for adaptation only and the nonlinear terms are approximated via interpolation once the DEIM interpolants have been adapted. Other sampling strategies motivated by DEIM and GappyPOD are investigated by Manohar et al. [29], who showed improvements for signal reconstruction [30, 38]. Greedy methods for sensor placement in the context of empirical interpolation are investigated in [1, 8].

We consider GappyPOD in the specific setting where samples are polluted with noise. Noise is here to be understood in general terms, including perturbations that are typically modeled via random noise such as in turbulence; see, e.g., [36]. It has been discussed in [2] that the L_2 error of (D)EIM approximations can grow with the dimension of the (D)EIM space in the presence of noise. The work [2] proposes taking

more sampling points than the dimension of the (D)EIM space as a possible remedy and demonstrates on numerical results that this gives more stable results than (D)EIM, i.e., that the error does not increase with the (D)EIM dimension. We build on the vast literature on GappyPOD and related methods [3, 4, 12, 13, 29, 43, 45]. Our contribution is a probabilistic analysis that proves that in expectation with high probability GappyPOD with oversampling avoids the increase of the L_2 error with the dimension of the reduced space. For the analysis, we follow the work by Balzano, Recht, and Nowak [5] and the work by Cohen, Davenport, and Leviatan [16] that provide approximation results for least-squares approximations, which we apply to GappyPOD with oversampling. Extensions to the work by Cohen, Davenport, and Leviatan [16] have been introduced in [15, 31]. We then discuss a deterministic oversampling strategy and demonstrate with numerical results that a lower error with GappyPOD is achieved in the presence of noise compared to (D)EIM that interpolates the nonlinear terms.

The structure of the paper is as follows. Section 2 briefly reviews DEIM in the context of model reduction and numerically demonstrates on a toy example that DEIM approximations are unstable if the nonlinear function evaluations are polluted with noise. Sections 3 and 4 analyze GappyPOD with randomized samples and prove that oversampling avoids the stability issue in expectation with high probability. Section 5 introduces a deterministic sampling strategy, which is then shown to achieve more accurate reduced models than (D)EIM in section 6.

2. Preliminaries and problem formulation.

This section briefly reviews (D)EIM for approximating the nonlinear terms in reduced models and for recovering field data from few measurements and demonstrates, via an example, that (D)EIM can become unstable in the presence of noise.

2.1. Model reduction with empirical interpolation.

Consider a system of parametrized nonlinear equations

$$\boldsymbol{A}\boldsymbol{x}(\boldsymbol{\xi}) + \boldsymbol{f}(\boldsymbol{x}(\boldsymbol{\xi}); \boldsymbol{\xi}) = 0\,, \tag{1}$$

where $\boldsymbol{x}(\boldsymbol{\xi}) \in \mathbb{R}^N$ is the state, $\boldsymbol{\xi} \in \mathcal{D}$ is a d-dimensional parameter in the parameter domain $\mathcal{D}$, $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is a constant matrix, and $\boldsymbol{f} : \mathbb{R}^N \times \mathcal{D} \to \mathbb{R}^N$ is a nonlinear function. Systems such as (1) typically arise after discretizing a PDE in the spatial domain, in which case the matrix $\boldsymbol{A}$ corresponds to the linear operators of the underlying PDE and the nonlinear function $\boldsymbol{f}$ to the nonlinear terms. In the following, we are interested in situations where the dimension $N \in \mathbb{N}$ of the state $\boldsymbol{x}(\boldsymbol{\xi})$ is large, which means that system (1) is potentially expensive to solve numerically, especially if these simulations need to be repeated for many parameter samples in outer-loop applications [35] such as optimization, uncertainty quantification, and control.

A common approach to constructing a reduced model of the full system (1) is to use projection-based model reduction [7, 37]. Toward this goal, let the columns of the matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_M] \in \mathbb{R}^{N\times M}$ be M snapshots derived from the parameter samples $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_M \in \mathcal{D}$ such that $\boldsymbol{x}_i = \boldsymbol{x}(\boldsymbol{\xi}_i)$ for $i = 1, \dots, M$. Note that typically $M \leq N$. Further, let $\boldsymbol{V} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_r] \in \mathbb{R}^{N\times r}$ be an r-dimensional orthonormal basis constructed from the snapshot matrix $\boldsymbol{X}$. A common approach to obtaining $\boldsymbol{V}$ is to compute the singular value decomposition (SVD) of $\boldsymbol{X}$ and then to define $\boldsymbol{V}$ as the leading $r \leq M$ left-singular vectors, as done in proper orthogonal decomposition (POD). Then, the POD-Galerkin reduced model is obtained via projection

$$\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}}(\boldsymbol{\xi}) + \boldsymbol{V}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi}); \boldsymbol{\xi}) = 0\,, \tag{2}$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{V}^T\boldsymbol{A}\boldsymbol{V}$ is the reduced linear operator and $\tilde{\boldsymbol{x}}(\boldsymbol{\xi}) \in \mathbb{R}^r$ is the reduced state.

Even though the reduced state $\tilde{\boldsymbol{x}}(\boldsymbol{\xi})$ is in the r-dimensional subspace, evaluation of the reduced nonlinear term $\boldsymbol{V}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})$ in (2) still requires, first, lifting $\tilde{\boldsymbol{x}}(\boldsymbol{\xi})$ to the full dimension N, evaluating the original nonlinear term in this original dimension, and then projecting it down to the reduced dimension; thus, evaluating the reduced model (2) still requires operations that scale with the dimension of the full model. This is called the lifting bottleneck in model reduction.

An effective remedy to the lifting bottleneck is empirical interpolation [6, 14]. The goal is to find an accurate approximation $\tilde{\boldsymbol{f}} : \mathbb{R}^n \times \mathcal{D} \to \mathbb{R}^n$ to $\boldsymbol{f}$ that is computationally cheap to evaluate with cost independent of the dimension N. The empirical interpolation approximant $\tilde{\boldsymbol{f}}$ has the form

$$\tilde{\boldsymbol{f}}\,(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = \boldsymbol{U}\boldsymbol{c}(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) \tag{3}$$

with $\tilde{\boldsymbol{x}}(\boldsymbol{\xi}) \in \mathbb{R}^n$ and where $\boldsymbol{c}(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) \in \mathbb{R}^n$ are the coefficients of the linear combination with the columns of $\boldsymbol{U} \in \mathbb{R}^{N\times n}$, which form a basis of an n-dimensional reduced space in which to approximate the function $\boldsymbol{f}$ with $n \ll N$. DEIM achieves the approximation (3) by interpolating $\boldsymbol{f}$ at selected components. Let $p_1, \ldots, p_n \in \{1, \ldots, N\}$ be the interpolation points (indices), i.e., $\boldsymbol{e}_{p_i}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = \boldsymbol{e}_{p_i}^T\tilde{\boldsymbol{f}}\,(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})$ for $i = 1, 2, \ldots, n$, where $\boldsymbol{e}_i \in \mathbb{R}^N$ denotes the ith canonical unit vector. Let $\boldsymbol{P} = [\boldsymbol{e}_{p_1}, \ldots, \boldsymbol{e}_{p_n}] \in \mathbb{R}^{N\times n}$ be the corresponding interpolation points (index selection) matrix. Then, the interpolation conditions are $\boldsymbol{P}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = \boldsymbol{P}^T\tilde{\boldsymbol{f}}\,(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})$, which, using (3), lead to

$$\tilde{\boldsymbol{f}}\,(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = \boldsymbol{U}\boldsymbol{c}(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})\,, \tag{4}$$

where $\boldsymbol{c}(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = (\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})$. In (4), $\tilde{\boldsymbol{f}}$ is the DEIM approximation of $\boldsymbol{f}$.

The columns of $\boldsymbol{U} \in \mathbb{R}^{N\times n}$ are often taken as the POD basis of the nonlinear snapshots $\boldsymbol{f}(\boldsymbol{x}(\boldsymbol{\xi}_1);\boldsymbol{\xi}_1), \ldots, \boldsymbol{f}(\boldsymbol{x}(\boldsymbol{\xi}_M);\boldsymbol{\xi}_M)$ with parameters $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M \in \mathcal{D}$. Note that $\boldsymbol{U}$ is orthonormal. The choice of the selection operator $\boldsymbol{P}$ is motivated by the error bound

$$\|\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) - \tilde{\boldsymbol{f}}\,(\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})\|_2 \leq \left\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\right\|_2 \left\|(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}^T)\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})\right\|_2, \tag{5}$$

where $\|(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}^T)\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})\|_2$ is the error due to the optimal approximation by orthogonal projection; see [6, 14]. Therefore, the selection operator $\boldsymbol{P}$ should choose indices such that $\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2$ is small. The DEIM algorithm [6, 14] performs a greedy search to select the interpolation points. The QDEIM point selection algorithm [17, 18] based on the rank-revealing QR factorization is an alternative to this greedy-based point selection algorithms. Combining the DEIM approximation (4) with the POD-Galerkin reduced model (2), we obtain the POD-DEIM-Galerkin reduced model

$$\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}}(\boldsymbol{\xi}) + \boldsymbol{V}^T\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi}) = 0\,, \tag{6}$$

where the $N - n$ components of $\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})$ that are different from the interpolation points $p_1, \ldots, p_n$ are approximated via empirical interpolation. Thus, the reduced model (6) requires evaluating the nonlinear function $\boldsymbol{f}$ at only n components, which typically leads to significant speedups compared to the POD-Galerkin reduced model (2) that requires evaluating the function $\boldsymbol{f}$ at all N components.

2.2. State field approximation from few measurements. Another use case of empirical interpolation and related methods, such as GappyPOD [20], is approximating state fields $\boldsymbol{x} : \mathcal{D} \to \mathbb{R}^N$ from a few spatial measurements [1, 2, 9, 42], where

$\boldsymbol{\xi} \in \mathcal{D}$ is a parameter that defines the field $\boldsymbol{x}(\boldsymbol{\xi})$. Let $\boldsymbol{U} \in \mathbb{R}^{N\times n}$ be the reduced basis matrix constructed from snapshots $\boldsymbol{x}(\boldsymbol{\xi}_1), \ldots, \boldsymbol{x}(\boldsymbol{\xi}_M)$ for $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M \in \mathcal{D}$ via, e.g., POD, and let $\boldsymbol{P} \in \mathbb{R}^{N\times n}$ be the interpolation points matrix derived from $\boldsymbol{U}$ with, e.g., the greedy algorithm [6, 14] and QDEIM [17, 18]. Given are the measurements $\boldsymbol{x}_P(\boldsymbol{\xi}) = \boldsymbol{P}^T\boldsymbol{x}(\boldsymbol{\xi})$ at the spatial coordinates corresponding to the components selected by $\boldsymbol{P}$ of a field $\boldsymbol{x}(\boldsymbol{\xi}) \in \mathbb{R}^N$ with parameter $\boldsymbol{\xi} \in \mathcal{D}$. The field $\boldsymbol{x}(\boldsymbol{\xi})$ is unknown at all spatial coordinates except at the interpolation points given by $\boldsymbol{P}$. The DEIM approximation of $\boldsymbol{x}(\boldsymbol{\xi})$ is then given by $\tilde{\boldsymbol{x}}(\boldsymbol{\xi}) = \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{x}_P(\boldsymbol{\xi})$. For the ease of presentation, we follow the notation introduced in subsection 2.1 for approximating nonlinear terms for model reduction; however, all of what is presented in the following directly applies to state field approximation as well. We will revisit state field approximation in our numerical experiments in section 6.

2.3. Instability of empirical interpolation in the presence of noise. To approximate $\boldsymbol{f}(\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi});\boldsymbol{\xi})$ with DEIM in the reduced model (6), the function $\boldsymbol{f}$ is evaluated (at least) at the components of $\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi})$ corresponding to the interpolation points $p_1, \ldots, p_n$, while all the other components are approximated via interpolation in the reduced space spanned by the columns of the basis matrix $\boldsymbol{U}$. We are interested in the situation where the function evaluations of $\boldsymbol{f}$ at $\boldsymbol{V}\tilde{\boldsymbol{x}}(\boldsymbol{\xi})$ are noisy, in which case DEIM approximations can become unstable, as demonstrated in, e.g., [2].

Consider the parametrized nonlinear function

$$\boldsymbol{f}(\boldsymbol{x};\xi) = \exp\left(-\frac{(\boldsymbol{x}-\xi)^2}{5\times 10^{-3}}\right) \tag{7}$$

with the parameter $\xi \in \mathcal{D} = [1,3] \subset \mathbb{R}$. The components of $\boldsymbol{x} \in \mathbb{R}^{8192}$ are the equidistant points in $\Omega = [-2\pi, 2\pi]$. Note that all operations in (7) are to be understood componentwise. Let $\xi_1, \ldots, \xi_{2500}$ be the equidistant points in $\mathcal{D}$ and let $\boldsymbol{f}(\boldsymbol{x};\xi_1), \ldots, \boldsymbol{f}(\boldsymbol{x};\xi_{2500})$ be the nonlinear snapshots to derive a DEIM interpolant $\tilde{\boldsymbol{f}}$ of $\boldsymbol{f}$ of dimension n with the reduced basis matrix $\boldsymbol{U}$ and the QDEIM interpolation points matrix $\boldsymbol{P}$. We now approximate the function $\boldsymbol{f}$ at the 2500 parameters $\xi'_1, \ldots, \xi'_{2500} \in \mathcal{D}$ uniformly sampled in the domain $\mathcal{D}$. Note that the parameters $\xi'_1, \ldots, \xi'_{2500}$ are different from the parameters $\xi_1, \ldots, \xi_{2500}$ that were used to construct the reduced space and the interpolation points matrix. The DEIM approximation $\tilde{\boldsymbol{f}}$ of $\boldsymbol{f}$ is

$$\tilde{\boldsymbol{f}}(\boldsymbol{x};\xi'_i) = \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}(\boldsymbol{x};\xi'_i)$$

for $i = 1, \ldots, 2500$. The averaged relative state error

$$\frac{1}{2500}\sum_{i=1}^{2500}\frac{\|\boldsymbol{f}(\boldsymbol{x};\xi'_i) - \tilde{\boldsymbol{f}}(\boldsymbol{x};\xi'_i)\|_2}{\|\boldsymbol{f}(\boldsymbol{x};\xi'_i)\|_2} \tag{8}$$

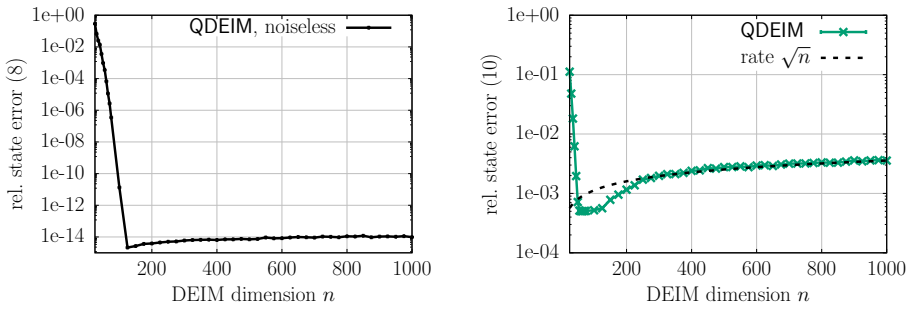
versus the dimension n of the DEIM approximation is plotted in Figure 1(a). The results indicate a fast decay of the DEIM approximation error with the dimension n.

Let us now consider noisy evaluations of the function $\boldsymbol{f}$. Therefore, let $\boldsymbol{\epsilon}$ be a random vector that has, as components, independent zero-mean Gaussian random variables with standard deviation $\sigma = 10^{-4}$. Define

$$\boldsymbol{f}_{\boldsymbol{\epsilon}}(\boldsymbol{x};\xi) = \boldsymbol{f}(\boldsymbol{x};\xi) + \boldsymbol{\epsilon} \tag{9}$$

so that the DEIM approximation using the noisy function evaluations (9) is

$$\tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}}(\boldsymbol{x};\xi) = \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}_{\boldsymbol{\epsilon}}(\boldsymbol{x};\xi)\,.$$



(a) without noise (b) with noise

FIG. 1. *The (D)EIM is sensitive to noise in the sparse samples of the nonlinear function. In particular, the noise is amplified as the dimension n of the reduced space is increased. A rate of* $\sqrt{n}$ *is numerically observed. Standard deviation of noise is* 10^{-4}.

The plot in Figure 1(b) shows the averaged relative state error

$$\sum_{j=1}^{k} \frac{1}{2500} \sum_{i=1}^{2500} \frac{\|\boldsymbol{f}(\boldsymbol{x};\xi_i') - \tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}_j}(\boldsymbol{x};\xi_i')\|_2}{\|\boldsymbol{f}(\boldsymbol{x};\xi_i')\|_2} \tag{10}$$

for $k = 10$ replicates of the DEIM approximation $\tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}_j}$ that is derived from the noisy function evaluations (9) with realization $\boldsymbol{\epsilon}_j$ of the noise. The error bars indicate the minimum and maximum of the error over the replicates. Note that the error bars are barely visible, which means the variation over the replicates is small. The results indicate a stability issue of DEIM in this case of noisy function evaluations because the error grows with the dimension n of the reduced space. The result illustrates an error growth with a rate $\sqrt{n}$ with the dimension n. Similar observations are made in [2].

Remark 2.1. The term "instability" has various meanings in numerical analysis. In the following, the term "instability" refers to the specific phenomenon that the DEIM approximation error $\|\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\xi}) - \tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}}(\boldsymbol{x};\boldsymbol{\xi})\|_2$ in the Euclidean norm grows with the dimension n of the reduced space if noisy function evaluations (9) (or noisy measurements in the context of state field approximation in subsection 2.2) are used; see Figure 1(b).

3. Amplification of noise in DEIM. We provide an upper bound on the amplification of the noise in DEIM approximations, and a theoretical explanation of the numerical observation in Figure 1(b). The bound (12), which we prove in the following, shows that the error cannot increase faster than with rate $\sqrt{n}$, which is the rate observed in Figure 1(b). We also provide a formula for the expected value of the DEIM error vector and reveal the structure of the error ellipsoid. A bound similar to (12) has been presented in [2].

To simplify the exposition, we drop the dependence on the state $\boldsymbol{x}$ and the parameter $\boldsymbol{\xi}$ of $\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\xi})$ and abbreviate it as $\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\xi}) = \boldsymbol{f}$. Similarly, the DEIM approximant will be abbreviated as $\tilde{\boldsymbol{f}}$. The noisy counterparts of $\boldsymbol{f}$ and $\tilde{\boldsymbol{f}}$ are $\boldsymbol{f}_{\boldsymbol{\epsilon}} = \boldsymbol{f} + \boldsymbol{\epsilon}$ and $\tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}}$, respectively, where $\boldsymbol{\epsilon}$ is a zero-mean Gaussian vector with independent components with standard deviation $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_N]^T$.

LEMMA 3.1. *Define the error of the DEIM approximation* $\tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}}$ *from noisy function evaluations as* $\boldsymbol{r}_{\boldsymbol{\epsilon}} = \boldsymbol{f} - \tilde{\boldsymbol{f}}_{\boldsymbol{\epsilon}} = \boldsymbol{f} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}_{\boldsymbol{\epsilon}}$ *and the error of the approxima-*

tion $\tilde{\boldsymbol{f}}$ with noise-free function evaluations as $\boldsymbol{r} = \boldsymbol{f} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}$. Then, the expected value of the error $\boldsymbol{r}_{\boldsymbol{\epsilon}}$ corresponding to noisy function evaluations equals $\boldsymbol{r}$, i.e., $\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}] = \boldsymbol{r}$, where the expectation is taken over the noise. The standard deviation of $\boldsymbol{r}_{\boldsymbol{\epsilon}}$ satisfies

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{r}_{\boldsymbol{\epsilon}} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2\right] \leq \sqrt{\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{r}_{\boldsymbol{\epsilon}} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2^2\right]} \leq \sqrt{n}\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2\|\boldsymbol{P}^T\boldsymbol{\sigma}\|_\infty\,. \tag{11}$$

Thus, the error is bounded in expectation as

$$\mathbb{E}_{\boldsymbol{\epsilon}}\big[\|\boldsymbol{f} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}_{\boldsymbol{\epsilon}}\|_2\big] \leq \|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2\big(\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2 + \sqrt{n}\|\boldsymbol{P}^T\boldsymbol{\sigma}\|_\infty\big)\,. \tag{12}$$

Proof. Using the linearity of the expectation, the error formula for the DEIM projection, and the assumptions on the noise, namely, $\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}] = 0$, we obtain

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\epsilon}}\left[\boldsymbol{f} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}_{\boldsymbol{\epsilon}}\right] &= (\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T)\boldsymbol{f} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}] \\ &= (\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T)\boldsymbol{f},\end{aligned} \tag{13}$$

which establishes $\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}] = \boldsymbol{r}$ as claimed. The norm of $\|\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2$ is bounded as

$$\left\|\mathbb{E}_{\boldsymbol{\epsilon}}\left[\boldsymbol{f} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{f}_{\boldsymbol{\epsilon}}\right]\right\|_2 \leq \left\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\right\|_2\left\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\right\|_2, \tag{14}$$

which is the same upper bound as in (5) for the noise-free case. The covariance matrix of the error $\boldsymbol{r}_{\boldsymbol{\epsilon}}$ is

$$\begin{aligned}\boldsymbol{C} = \mathbb{E}_{\boldsymbol{\epsilon}}\left[\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\boldsymbol{P}(\boldsymbol{P}^T\boldsymbol{U})^{-T}\boldsymbol{U}^T\right] &= \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\boldsymbol{P}^T\Sigma^2\boldsymbol{P}(\boldsymbol{P}^T\boldsymbol{U})^{-T}\boldsymbol{U}^T \\ &= \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^{-1}\Sigma_{\boldsymbol{P}}^2(\boldsymbol{P}^T\boldsymbol{U})^{-T}\boldsymbol{U}^T\,,\end{aligned} \tag{15}$$

where $\Sigma_{\boldsymbol{P}}^2 = \boldsymbol{P}^T\Sigma^2\boldsymbol{P} = \operatorname{diag}(\sigma_{p_i}^2)_{i=1}^n$ with σ_{p_i} being the standard deviation of the p_ith component of $\boldsymbol{\epsilon}$. The covariance $\boldsymbol{C}$ is positive semidefinite. Its nonzero eigenvalues λ_i^2 (that correspond to the invariant space spanned by $\boldsymbol{U}$) can be enumerated so that

$$\lambda_i^2 = \sigma_{p_i}^2\vartheta_i^2, \text{ where } \frac{1}{\|\boldsymbol{P}^T\boldsymbol{U}\|_2} \leq \vartheta_i \leq \|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2, \quad i = 1, \ldots, n.$$

This is an application of the Ostrowski theorem [24, Theorem 4.5.9]; it identifies the bounds of the amplification factors ϑ_i's of the corresponding standard deviations. The spectral structure of $\boldsymbol{C}$ (and thus the error ellipsoid) can be explicitly revealed using the SVD $\boldsymbol{\Sigma}_{\boldsymbol{P}}^{-1}(\boldsymbol{P}^T\boldsymbol{U}) = \boldsymbol{\Phi}\boldsymbol{\Omega}\boldsymbol{\Psi}^T$ ($\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ orthogonal matrices of singular vectors, $\boldsymbol{\Omega}$ diagonal matrix of singular values), which yields $\boldsymbol{C} = (\boldsymbol{U}\boldsymbol{\Psi})\boldsymbol{\Omega}^{-2}(\boldsymbol{U}\boldsymbol{\Psi})^T$.

Next, of interest is the variance $\mathbb{E}_{\boldsymbol{\epsilon}}[\|\boldsymbol{r}_{\boldsymbol{\epsilon}} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2^2]$ of $\boldsymbol{r}_{\boldsymbol{\epsilon}}$, for which it follows that

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{r}_{\boldsymbol{\epsilon}} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2^2\right] = \operatorname{Trace}(\boldsymbol{C}) = \sum_{i=1}^n \sigma_{p_i}^2\vartheta_i^2 \leq n\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2^2 \max_i \sigma_{p_i}^2\,,$$

which shows $\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{r}_{\boldsymbol{\epsilon}} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2^2\right] \leq n\|\boldsymbol{P}^T\boldsymbol{\sigma}\|_\infty^2\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2^2$. In addition, by deploying Jensen's inequality and taking square root, we obtain (11). Then, combining the above estimates and the triangle inequality yields

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{r}_{\boldsymbol{\epsilon}}\|_2\right] &\leq \mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{r}_{\boldsymbol{\epsilon}} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{r}_{\boldsymbol{\epsilon}}]\|_2\right] + \|\mathbb{E}_{\boldsymbol{\epsilon}}\left[\boldsymbol{r}_{\boldsymbol{\epsilon}}\right]\|_2 \\ &\leq \|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2(\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2 + \sqrt{n}\|\boldsymbol{P}^T\boldsymbol{\sigma}\|_\infty),\end{aligned}$$

which proves (12). ∎

Remark 3.2. Lemma 3.1 assumes that the noise $\boldsymbol{\epsilon}$ is a zero-mean Gaussian vector with independent components. If $\boldsymbol{\epsilon}$ does not have independent components, i.e., if the noise error covariance matrix $\boldsymbol{\Sigma}^2$ is not diagonal, we can write the spectral decomposition of its submatrix $\boldsymbol{\Sigma}_P^2 = \boldsymbol{P}^T\boldsymbol{\Sigma}^2\boldsymbol{P}$ as $\boldsymbol{\Sigma}_P^2 = \boldsymbol{W_P}\boldsymbol{D_P^2}\boldsymbol{W_P}^T$, where $\boldsymbol{D_P} = \mathrm{diag}(\boldsymbol{d}_i)_{i=1}^n$. Using this spectral decomposition, we replace (15) with

$$\boldsymbol{C} = \boldsymbol{U}(\widetilde{\boldsymbol{P}}^T\boldsymbol{U})^{-1}\boldsymbol{D_P^2}(\widetilde{\boldsymbol{P}}^T\boldsymbol{U})^{-T}\boldsymbol{U}^T,$$

where $\widetilde{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{W_P}$. Since $\boldsymbol{W_P}$ is an orthogonal matrix, $\|\widetilde{\boldsymbol{P}}^T\boldsymbol{U}\|_2 = \|\boldsymbol{P}^T\boldsymbol{U}\|_2$ and $\|(\widetilde{\boldsymbol{P}}^T\boldsymbol{U})^{-1}\|_2 = \|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2$. Then, the rest of the proof of Lemma 3.1 follows as before by replacing $\boldsymbol{P}$ with $\widetilde{\boldsymbol{P}}$, and σ_{p_i} with $\boldsymbol{d}_i$ for $i = 1, \ldots, n$. Finally, the upper bounds (11) and (12) hold true by replacing $\boldsymbol{P}^T\boldsymbol{\sigma}$ with the vector $[\,\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n]^T$, where now $\boldsymbol{d}_i^2$'s are the variances along the principal components (eigenvectors of $\boldsymbol{\Sigma_P}$). Moreover, in both (11) and (12), the term $\sqrt{n}\|\boldsymbol{P}^T\boldsymbol{\sigma}\|_\infty$ can be replaced with $\|\boldsymbol{\Sigma_P}\|_F = \sqrt{\boldsymbol{d}_1^2 + \cdots + \boldsymbol{d}_n^2}$.

COROLLARY 3.3. *If the selection operator $\boldsymbol{P}$ is based on the quasi-optimal point selection introduced in* [18, Lemma 2.1]*, then* (11) *and* (12) *hold with the bound*

$$\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2 \leq \sqrt{1 + \eta^2 n(N-n)}, \tag{16}$$

where $\eta \geq 1$ is a tuning parameter.

Remark 3.4. Note how in (12), with increasing column dimension n of the matrix $\boldsymbol{U}$, the POD projection error $\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2$ monotonically decreases toward zero and, at the same time, the norm of the sampling operator $\|(\boldsymbol{P}^T\boldsymbol{U})^{-1}\|_2$ approaches one, while the contribution of the noise grows as $\sqrt{n}\|\boldsymbol{P}^T\boldsymbol{\sigma}\|_\infty$, taking over the leading term. The effect of the noise dominating the error is seen in Figure 1.

Remark 3.5. From Lemma 3.1, it follows that it is desirable that a DEIM selection operator avoids components of $\boldsymbol{f}$ with noise with high variance. Such a strategy may help slow the noise buildup. If we denote by $\mathcal{J}$ undesirable indices and set $\mathcal{J}^c = \{1, \ldots, N\}\backslash\mathcal{J}$, then we can run the QDEIM selection on the submatrix $\boldsymbol{U}(\mathcal{J}^c, :)$; for details we refer the reader to [17, section 3.].

4. Stability of GappyPOD with randomized samples.

Given the reduced basis matrix $\boldsymbol{U} \in \mathbb{R}^{N\times n}$, the DEIM selects n interpolation points, i.e., $\boldsymbol{P}^T\boldsymbol{U}$ is a square matrix. In this section, we investigate oversampling in the sense that more sampling points $m > n$ than the dimension n of the reduced space spanned by the columns of $\boldsymbol{U}$ are used. Taking more sampling points than the dimension of the space goes by the name of gappy proper orthogonal decomposition, GappyPOD, which was introduced in [20] and is used in the context for model reduction in [3, 4]. We now show that with GappyPOD, the noise amplification that was observed in section 3 can be avoided in expectation with high probability if sampling points are randomized and if more sampling points than basis vectors are used.

4.1. GappyPOD.

Consider $p_1, \ldots, p_m \in \{1, \ldots, N\}$, pairwise distinct sampling points with $m > n$, i.e., the number of sampling points m is larger than the dimension n of the space spanned by the columns of the basis matrix $\boldsymbol{U}$. Then, the corresponding GappyPOD approximation of $\boldsymbol{f}$ is

$$\hat{\boldsymbol{f}} = \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{f}\,,$$

where $\boldsymbol{M}^\dagger$ denotes the Moore–Penrose inverse of $\boldsymbol{M}$, i.e., $\boldsymbol{M}^\dagger = \left(\boldsymbol{M}^T\boldsymbol{M}\right)^{-1}\boldsymbol{M}^T$, assuming $\boldsymbol{M}$ has linearly independent columns. In contrast to the DEIM approximation

$\tilde{\boldsymbol{f}}$ in (4), the GappyPOD approximation $\hat{\boldsymbol{f}}$ is obtained via regression and therefore does not necessarily interpolate $\boldsymbol{f}$ at the sampling points $p_1, \dots, p_m$. In the case of noise-free sampling, the error of the GappyPOD approximation in the Euclidean norm satisfies (see, e.g., [45, Proposition 2.1])

$$\|\boldsymbol{f} - \hat{\boldsymbol{f}}\|_2 \leq \|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2 \|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2\,, \tag{17}$$

where $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2$ quantifies the effect of the sampling points and $\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2$ relates to the approximation quality of the space spanned by $\boldsymbol{U}$; cf. the DEIM error bound (5).

4.2. Probabilistic analysis of GappyPOD. We now investigate the error of the GappyPOD approximation $\hat{\boldsymbol{f}}$ when the sampling points $p_1, \dots, p_m$ are selected uniformly with replacement from $\{1, \dots, N\}$. Note that the following analysis is developed for uniform sampling *with replacement* as in the work by Balzano, Recht, and Nowak [5]. Parts of the following analysis are an application of the work by Cohen, Davenport, and Leviatan [16].

To set up the analysis, we define the coherence of a subspace $\mathcal{U} = \operatorname{span}(\boldsymbol{U})$ as

$$\mu(\mathcal{U}) = \frac{N}{n} \max_{i=1,\dots,N} \|\boldsymbol{u}_i^T\|_2^2\,,$$

where the columns of $\boldsymbol{U} \in \mathbb{R}^{N\times n}$ form an orthonormal basis for $\mathcal{U}$ and $\boldsymbol{u}_i^T$ is the ith row of $\boldsymbol{U}$; see, e.g., [10, Definition 1.2]. Intuitively speaking, coherence measures if there are certain coordinate directions that carry significantly more information than other directions. Note that $\max_{i=1,\dots,N} \|\boldsymbol{u}_i^T\|_2^2 \geq n/N$. The following result from [5, Lemma 3] will be used in our analysis.

LEMMA 4.1. *Let the points $p_1, \dots, p_m$ be uniformly sampled from $\{1, \dots, N\}$ with replacement and let $\boldsymbol{P}$ be the corresponding sampling points matrix. Moreover, let $\delta \in (0, 1]$ such that $m \geq (8/3)n\mu(\mathcal{U})\log(2n/\delta)$ and set $\gamma = \sqrt{\frac{8n\mu(\mathcal{U})}{3m}\log(2n/\delta)}$. Then*

$$\left\| \left(\left(\boldsymbol{P}^T\boldsymbol{U}\right)^T \left(\boldsymbol{P}^T\boldsymbol{U}\right) \right)^{-1} \right\|_2 \leq \frac{N}{(1-\gamma)m}$$

with probability at least $1 - \delta$.

The following lemma states that $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2$ can be bounded by a constant, arbitrarily close to 1, with high probability if a sufficiently large number of sampling points is used, which means that GappyPOD with randomized samples is well-posed with high probability.

LEMMA 4.2. *Consider the same setup as in Lemma 4.1 and set $\widehat{\gamma} = \sqrt{m}\gamma$. If m is such that, for $K \geq 1$,*

$$\sqrt{m} \geq \frac{1}{2}\widehat{\gamma} + \frac{1}{2}\sqrt{\widehat{\gamma}^2 + \frac{4N}{K^2}}, \tag{18}$$

then

$$\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2 \leq K$$

with probability at least $1 - \delta$.

Proof. Since

$$\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2 = \frac{1}{\sigma_{\min}(\boldsymbol{P}^T\boldsymbol{U})} = \sqrt{\frac{1}{\lambda_{\min}((\boldsymbol{P}^T\boldsymbol{U})^T(\boldsymbol{P}^T\boldsymbol{U}))}} = \sqrt{\|((\boldsymbol{P}^T\boldsymbol{U})^T(\boldsymbol{P}^T\boldsymbol{U}))^{-1}\|_2}$$

holds, Lemma 4.1 yields that with probability at least $1-\delta$

$$\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2 \leq \sqrt{\frac{N}{(1-\gamma)m}}. \tag{19}$$

Then, the task is to choose m so that $N/((1-\gamma)m) \leq K^2$. To that end, set $\widehat{\gamma} = \sqrt{m}\gamma$. By the assumption of Lemma 4.1, $\gamma \leq 1$ and thus $\sqrt{m} \geq \widehat{\gamma}$. The desired inequality becomes

$$N \leq K^2(m - \sqrt{m}\widehat{\gamma}), \text{ i.e., } K^2x^2 - K^2\widehat{\gamma}x - N \geq 0, \text{ where } x = \sqrt{m} \geq \widehat{\gamma}.$$

The smaller root of the above parabola is negative and the larger one, then, provides the desired lower bound (18). ∎

The following bound will be helpful in establishing the main result.

LEMMA 4.3. *Consider the same setup as in Lemma* 4.1. *Let* $\boldsymbol{g} \in \mathbb{R}^N$, *and let* α *be the acute angle between* $\boldsymbol{g}$ *and the range of* $\boldsymbol{U}$. *Then,*

$$\mathbb{E}_P\left[\left\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right] \leq \min\left\{\frac{1}{\sqrt{1-\gamma}}, \frac{\sqrt{\cos^2\alpha + \frac{n}{m}\mu(\mathcal{U})}}{1-\gamma}\right\}\|\boldsymbol{g}\|_2 \tag{20}$$

with probability at least $1-\delta$, *where the expected value* $\mathbb{E}_P$ *is with respect to the uniform distribution of the sampling points.*

Proof. We first apply submultiplicativity to obtain

$$\left\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\right\|_2 \leq \|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2\|\boldsymbol{P}^T\boldsymbol{g}\|_2\,. \tag{21}$$

Using (19) and Lemma 4.1, we have, with probability at least $1-\delta$, that

$$\mathbb{E}_P\left[\left\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right] \leq \sqrt{\frac{N}{(1-\gamma)m}}\;\mathbb{E}_P\left[\left\|\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right]. \tag{22}$$

Consider now the expected value $\mathbb{E}_P[\|\boldsymbol{P}^T\boldsymbol{g}\|_2^2]$ and note that we use the squared Euclidean norm $\|\cdot\|_2^2$. Let g_j denote the jth component of $\boldsymbol{g}$ for $j = 1, \ldots, N$. Also let $\mathbb{I}_{p_i=j}$ denote the indicator function that is 1 if $p_i = j$ and 0 otherwise. Note that the probability that $p_i = j$ is $1/N$ because a uniform distribution with replacement is used for selecting the sampling points, and thus $\mathbb{E}[\mathbb{I}_{p_i=j}] = 1/N$. Then,

$$\mathbb{E}_P\left[\left\|\boldsymbol{P}^T\boldsymbol{g}\right\|_2^2\right] = \mathbb{E}_P\left[\sum_{i=1}^{m}\sum_{j=1}^{N} g_j^2\,\mathbb{I}_{p_i=j}\right] = \frac{m}{N}\|\boldsymbol{g}\|_2^2\,. \tag{23}$$

Applying Jensen's inequality to (23) yields $\mathbb{E}_P\left[\left\|\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right] \leq \sqrt{\frac{m}{N}}\|\boldsymbol{g}\|_2$, which, combined with (22), implies

$$\mathbb{E}_P\left[\left\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right] \leq \frac{1}{\sqrt{1-\gamma}}\|\boldsymbol{g}\|_2, \tag{24}$$

proving the upper bound in (20) for the first input of the min function. To prove (20) for the second input of the min function, we first apply submultiplicativity to obtain

$$\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger \boldsymbol{P}^T\boldsymbol{g}\|_2 \leq \left\|\left((\boldsymbol{P}^T\boldsymbol{U})^T(\boldsymbol{P}^T\boldsymbol{U})\right)^{-1}\right\|_2 \|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\|_2\,. \tag{25}$$

With Lemma 4.1, we have, with probability at least $1-\delta$, that

$$\begin{aligned}\mathbb{E}_P\left[\left\|\left((\boldsymbol{P}^T\boldsymbol{U})^T(\boldsymbol{P}^T\boldsymbol{U})\right)^{-1}\right\|_2 \left\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right] \\ \leq \frac{N}{(1-\gamma)m}\mathbb{E}_P\left[\left\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right]\,.\end{aligned} \tag{26}$$

Let $\langle \boldsymbol{v}\,,\boldsymbol{w}\rangle_2 = \boldsymbol{v}^T\boldsymbol{w}$ denote the Euclidean inner product and consider the expected value $\mathbb{E}_P[\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\|_2^2]$. Using the linearity of $\langle\cdot,\cdot\rangle_2$ and $\mathbb{E}_P[\cdot]$, we obtain

$$\begin{aligned}\mathbb{E}_P\left[\left\langle(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g},(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\right\rangle_2\right] &= \mathbb{E}_P\left[\sum_{j=1}^m\sum_{k=1}^m\left\langle g_{p_j}\boldsymbol{u}_{p_j}^T, g_{p_k}\boldsymbol{u}_{p_k}^T\right\rangle_2\right] \\ &= \mathbb{E}_P\left[\sum_{j=1}^m\sum_{k=1}^m\left\langle\sum_{\ell=1}^N g_\ell\boldsymbol{u}_\ell^T\mathbb{I}_{p_j=\ell}, \sum_{s=1}^N g_s\boldsymbol{u}_s^T\mathbb{I}_{p_k=s}\right\rangle_2\right] \\ &= \sum_{j=1}^m\sum_{k=1}^m\sum_{\ell=1}^N\sum_{s=1}^N g_\ell g_s\left\langle\boldsymbol{u}_\ell^T,\boldsymbol{u}_s^T\right\rangle_2\mathbb{E}_P\left[\mathbb{I}_{p_j=\ell}\mathbb{I}_{p_k=s}\right]\,.\end{aligned} \tag{27}$$

Note that, for $k\neq j$, by independence of the jth and the kth drawing with replacement

$$\mathbb{E}_P\left[\mathbb{I}_{p_j=\ell}\mathbb{I}_{p_k=s}\right] = \frac{1}{N^2}\,. \tag{28}$$

Now we can use (28) to split (27). For $k\neq j$, we obtain

$$\sum_{j=1}^m\sum_{\substack{k=1\\k\neq j}}^m\sum_{\ell=1}^N\sum_{s=1}^N g_\ell g_s\left\langle\boldsymbol{u}_\ell^T,\boldsymbol{u}_s^T\right\rangle_2\frac{1}{N^2} = \frac{m^2-m}{N^2}\langle\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{g},\boldsymbol{g}\rangle_2 = \frac{m^2-m}{N^2}\|\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{g}\|_2^2.$$

The remaining terms with $j=k$ contribute to (27) with

$$\begin{aligned}\sum_{j=1}^m\sum_{\ell=1}^N\sum_{s=1}^N g_\ell g_s\left\langle\boldsymbol{u}_\ell^T,\boldsymbol{u}_s^T\right\rangle_2\mathbb{E}_P\left[\mathbb{I}_{p_j=\ell}\,\mathbb{I}_{p_j=s}\right] &= \frac{1}{N}\left[\sum_{j=1}^m\sum_{\ell=1}^N g_\ell^2\|\boldsymbol{u}_\ell^T\|_2^2\right] \\ &= \frac{m}{N}\sum_{\ell=1}^N g_\ell^2\|\boldsymbol{u}_\ell^T\|_2^2\,,\end{aligned}$$

where we use $\mathbb{E}_P[\mathbb{I}_{p_j=\ell}\mathbb{I}_{p_j=s}] = 1/N$ for $s=\ell$ and 0 otherwise. Altogether, we have

$$\begin{aligned}\mathbb{E}_P\left[\left\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\right\|_2^2\right] &= \frac{m^2-m}{N^2}\|\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{g}\|_2^2 + \frac{m}{N}\|\mathrm{diag}(\boldsymbol{g})\boldsymbol{U}\|_F^2 \\ &\leq \frac{m^2-m}{N^2}\|\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{g}\|_2^2 + \frac{m}{N}\frac{n}{N}\mu(\mathcal{U})\|\boldsymbol{g}\|_2^2 \qquad \left(\text{here } \frac{n}{N}\mu(\mathcal{U}) = \max_{\ell=1,\ldots,N}\|\boldsymbol{u}_\ell^T\|_2^2\right) \\ &= \frac{m^2-m}{N^2}\|\boldsymbol{g}\|_2^2\cos^2\alpha + \frac{m}{N}\frac{n}{N}\mu(\mathcal{U})\|\boldsymbol{g}\|_2^2\,,\end{aligned} \tag{29}$$

where $\cos^2\alpha = \|\boldsymbol{U}^T\boldsymbol{g}\|_2^2/\|\boldsymbol{g}\|_2^2$ and $\mathrm{diag}(\boldsymbol{g}) \in \mathbb{R}^{N\times N}$ is the diagonal matrix with the components of $\boldsymbol{g}$ on its diagonal. Hence, applying Jensen's inequality to (29) and then using (26) yields

$$\mathbb{E}_P\left[\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\|_2\right] \leq \frac{\|\boldsymbol{g}\|_2}{1-\gamma}\sqrt{\cos^2\alpha+\frac{n}{m}\mu(\mathcal{U})} = \frac{\|\boldsymbol{g}\|_2}{1-\gamma}\sqrt{\cos^2\alpha+\frac{N}{m}\max_{\ell=1,\ldots,N}\|\boldsymbol{u}_\ell^T\|_2^2}.$$

Combining this final inequality with (24) yields the desired result (20). ∎

Remark 4.4. The min function in the upper bound (20) results from using two different upper bounds for $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\|_2$, one as in (21) and the other one as in (25). The latter is employed by [5] for the special case where $\boldsymbol{g}$ is orthogonal to the range of $\boldsymbol{U}$. A technical calculation shows that while the first input in the min function is expected to be the smaller of the two for small m, the second input is expected to be the smaller one for large m.

Remark 4.5. If $\boldsymbol{g}$ is orthogonal to the range of $\boldsymbol{U}$, in (20) the term $\cos\alpha = 0$ and the upper bound simplifies to

$$\mathbb{E}_P\left[\left\|\left(\boldsymbol{P}^T\boldsymbol{U}\right)^\dagger\boldsymbol{P}^T\boldsymbol{g}\right\|_2\right] \leq \min\left\{\frac{1}{\sqrt{1-\gamma}}, \frac{\sqrt{\frac{n}{m}\mu(\mathcal{U})}}{1-\gamma}\right\}\|\boldsymbol{g}\|_2.$$

Thus, Lemma 4.3 contains [5, Lemma 2] as a special case. Indeed, in this special case, the expectation of $(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}$ is zero because

$$\mathbb{E}_P\left[(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{g}\right] = \mathbb{E}_P\left[\sum_{k=1}^{m}\sum_{j=1}^{N}\boldsymbol{u}_j^T g_j \mathbb{I}_{p_k=j}\right] = \frac{m}{N}\boldsymbol{U}^T\boldsymbol{g} = 0.$$

We now show that GappyPOD is robust with respect to noise in the sense that increasing the number of sampling points m reduces the effect of the noise.

THEOREM 4.6. *Consider the same setup as in Lemmas* 4.1 *and* 4.3. *Define*

$$\zeta = \min\left\{\frac{1}{\sqrt{1-\gamma}}, \frac{1}{1-\gamma}\sqrt{\frac{n}{m}\mu(\mathcal{U})}\right\}. \tag{30}$$

Then,

$$\mathbb{E}_P\left[\mathbb{E}_\epsilon\left[\|\boldsymbol{f}-\hat{\boldsymbol{f}}_{\boldsymbol{\epsilon}}\|_2\right]\right] \leq (1+\zeta)\|\boldsymbol{f}-\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2 + \frac{\|\boldsymbol{\sigma}\|_\infty}{(1-\gamma)}\sqrt{\frac{nN}{m}} \tag{31}$$

with probability at least $1-\delta$, *where the expectation* $\mathbb{E}_P$ *is with respect to the distribution of the samples and* $\mathbb{E}_\epsilon$ *with respect to the noise in* $\boldsymbol{f}_{\boldsymbol{\epsilon}}$ *as defined in* (9).

Proof. We split the error following the strategy of [16, Theorem 2]. With the triangular inequality, we obtain

$$\mathbb{E}_P\left[\mathbb{E}_\epsilon\left[\|\boldsymbol{f}-\hat{\boldsymbol{f}}_{\boldsymbol{\epsilon}}\|_2\right]\right] \leq \mathbb{E}_P\left[\|\boldsymbol{f}-\hat{\boldsymbol{f}}\|_2\right] + \mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{\epsilon}\|_2\right]\right] \tag{32}$$

with $\hat{\boldsymbol{f}} = \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{f}$ and $\boldsymbol{f}_{\boldsymbol{\epsilon}} = \boldsymbol{f}+\boldsymbol{\epsilon}$.

To bound the first term on the right-hand side of the inequality (32), set $\boldsymbol{g} = \boldsymbol{f}-\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}$. Similarly to the case in [14], it holds that

$$\boldsymbol{f}-\hat{\boldsymbol{f}} = \boldsymbol{f}-\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{f} = \boldsymbol{g}-\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\,, \tag{33}$$

where we used the fact that $(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{U} = \boldsymbol{I}_n$ with probability, at least, $1-\delta$. Note that $\boldsymbol{g} = \boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}$ is orthogonal to the range of $\boldsymbol{U}$. Then, Lemma 4.3 implies that

$$\mathbb{E}_P[\|\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{g}\|_2] \leq \min\left\{\frac{1}{\sqrt{1-\gamma}}, \frac{\sqrt{\frac{n}{m}\mu(\mathcal{U})}}{1-\gamma}\right\}\|\boldsymbol{g}\|_2 \tag{34}$$

with probability at least $1-\delta$. Then, (33) and the linearity of the expectation yield

$$\mathbb{E}_P\left[\|\boldsymbol{f} - \hat{\boldsymbol{f}}\|_2\right] \leq (1+\zeta)\,\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2, \tag{35}$$

where ζ is as defined in (30). Now consider the second term on the right-hand side of (32). Note that $\boldsymbol{\epsilon}$ is not necessarily orthogonal to $\boldsymbol{U}$, and therefore Remark 4.5 cannot be applied. We make the approximations

$$\mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{\epsilon}\|_2\right]\right] \leq \frac{N}{(1-\gamma)m}\mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{\epsilon}\|_2\right]\right], \tag{36}$$

which holds with probability at least $1-\delta$; see (25) and (26) in the proof of Lemma 4.3. Consider now $\mathbb{E}_P[\mathbb{E}_{\boldsymbol{\epsilon}}[\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{\epsilon}\|_2^2]]$. With the same notation as in the proof of Lemma 4.3, and building on the proof of [16, Theorem 3], we have

$$\begin{aligned}
&\mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{\epsilon}\|_2^2\right]\right] = \mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\left\langle(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{\epsilon}, (\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{\epsilon}\right\rangle_2\right]\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\left\langle\sum_{i=1}^m \epsilon_{p_i}\boldsymbol{u}_{p_i}^T, \sum_{j=1}^m \epsilon_{p_j}\boldsymbol{u}_{p_j}^T\right\rangle_2\right]\right] = \mathbb{E}_P\left[\sum_{i=1}^m\sum_{j=1}^m \mathbb{E}_{\boldsymbol{\epsilon}}\left[\left\langle\epsilon_{p_i}\boldsymbol{u}_{p_i}^T, \epsilon_{p_j}\boldsymbol{u}_{p_j}^T\right\rangle_2\right]\right] \\
&= \mathbb{E}_P\left[\sum_{i=1}^m\sum_{j=1}^m \mathbb{E}_{\boldsymbol{\epsilon}}\left[\epsilon_{p_i}\epsilon_{p_j}\left\langle\boldsymbol{u}_{p_i}^T, \boldsymbol{u}_{p_j}^T\right\rangle_2\right]\right] = \mathbb{E}_P\left[\sum_{i=1}^m\sum_{j=1}^m \mathbb{E}_{\boldsymbol{\epsilon}}\left[\epsilon_{p_i}\epsilon_{p_j}\right]\left\langle\boldsymbol{u}_{p_i}^T, \boldsymbol{u}_{p_j}^T\right\rangle_2\right] \\
&= \mathbb{E}_P\left[\sum_{i=1}^m \mathbb{E}_{\boldsymbol{\epsilon}}\left[\epsilon_{p_i}^2\right]\left\langle\boldsymbol{u}_{p_i}^T, \boldsymbol{u}_{p_i}^T\right\rangle_2\right] \quad \text{(since } \epsilon_{p_i} \text{ and } \epsilon_{p_j} \text{ are independent for } i \neq j\text{)} \\
&= \mathbb{E}_P\left[\sum_{i=1}^m \sigma_{p_i}^2\left\langle\boldsymbol{u}_{p_i}^T, \boldsymbol{u}_{p_i}^T\right\rangle_2\right] = \mathbb{E}_P\left[\sum_{i=1}^m\sum_{j=1}^N \sigma_j^2\|\boldsymbol{u}_j^T\|_2^2\mathbb{I}_{p_i=j}\right] = \frac{m}{N}\sum_{j=1}^N \sigma_j^2\|\boldsymbol{u}_j^T\|_2^2\,.
\end{aligned}$$

Using the fact that $\sum_{j=1}^N \|\boldsymbol{u}_j^T\|_2^2 = \|\boldsymbol{U}\|_F^2 = n$, we obtain

$$\mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\left\|(\boldsymbol{P}^T\boldsymbol{U})^T\boldsymbol{P}^T\boldsymbol{\epsilon}\right\|_2^2\right]\right] = \frac{m}{N}\sum_{j=1}^N \sigma_j^2\|\boldsymbol{u}_j^T\|_2^2 \leq \frac{m}{N}\|\boldsymbol{\sigma}\|_\infty^2\sum_{j=1}^N \|\boldsymbol{u}_j^T\|_2^2 = \frac{mn}{N}\|\boldsymbol{\sigma}\|_\infty^2.$$

Applying Jensen's inequality together with (36) yields

$$\mathbb{E}_P\left[\mathbb{E}_{\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger\boldsymbol{P}^T\boldsymbol{\epsilon}\right\|_2\right]\right] \leq \frac{N}{(1-\gamma)m}\sqrt{\frac{mn}{N}\|\boldsymbol{\sigma}\|_\infty^2} = \frac{\|\boldsymbol{\sigma}\|_\infty}{(1-\gamma)}\sqrt{\frac{nN}{m}}$$

with probability at least $1-\delta$. Combining this with (35) proves the theorem. ∎

Remark 4.7. Theorem 4.6 reveals that as $m \to \infty$ (recall we perform uniform sampling with replacement), the upper bound in (31) converges to the projection error $\|\boldsymbol{f} - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}\|_2$. In the numerical results in section 6, rather than investigating the asymptotic behavior as $m \to \infty$, we will typically keep the ratio n/m low by increasing m with n and so preventing that the noise term in (31) dominates.

5. The GappyPOD+E deterministic sampling strategy. In this section, we present a deterministic strategy that selects sampling points to reduce the quantity $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2$, which controls how sensitive the GappyPOD oblique projection is to perturbations and noise; cf. subsection 2.3. While our probabilistic analysis in section 4 shows that sampling points that are selected uniformly in $\{1, \ldots, N\}$ are sufficient for GappyPOD to be robust with respect to noise, the number of uniformly selected sampling points that are required grows with, e.g., the coherence $\mu(\mathcal{U})$ of the space $\mathcal{U}$. The following deterministic selection strategy aims to achieve robustness with fewer points than uniform sampling by taking information about the space $\mathcal{U}$ into account. We refer to the introduction for references to other sampling strategies.

We propose the GappyPOD+E sampling algorithm that is based on lower bounds of the smallest eigenvalues of certain structured matrix updates introduced in [26] and that is a special case of the approach introduced by Zimmermann and Willcox in [45]. The "E" in GappyPOD+E stands for "eigenvector." The goal of the GappyPOD+E sampling algorithm is to select points that minimize $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2$. This minimization problem is equivalent to maximizing the smallest singular value of $\boldsymbol{P}^T\boldsymbol{U}$ because

$$\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2 = s_{\max}\left((\boldsymbol{P}^T\boldsymbol{U})^\dagger\right) = \frac{1}{s_{\min}(\boldsymbol{P}^T\boldsymbol{U})}\,,$$

where $s_{\max}(\boldsymbol{M})$ and $s_{\min}(\boldsymbol{M})$ denote the largest and the smallest singular value of the matrix $\boldsymbol{M}$, respectively. The GappyPOD+E algorithm relies on lower bounds of the smallest eigenvalues to select points that maximize $s_{\min}(\boldsymbol{P}^T\boldsymbol{U})$ by leveraging the eigenvector corresponding to the smallest eigenvalue.

5.1. Singular values after symmetric rank-one updates. Consider the basis matrix $\boldsymbol{U}$ and the sampling points matrix[1] $\boldsymbol{P}_m$ that takes $m \geq n$ samples. Consider now the SVD of $\boldsymbol{P}_m^T\boldsymbol{U} \in \mathbb{R}^{m\times n}$

$$\boldsymbol{V}_m\boldsymbol{\Sigma}_m\boldsymbol{W}_m^T = \boldsymbol{P}_m^T\boldsymbol{U}\,,$$

where $\boldsymbol{V}_m \in \mathbb{R}^{m\times n}$ contains, as its columns, the left-singular vectors, the matrix $\boldsymbol{\Sigma}_m = \mathrm{diag}[s_1^{(m)}, \ldots, s_n^{(m)}] \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the singular values $s_1^{(m)}, \ldots, s_n^{(m)}$, in descending order, and $\boldsymbol{W}_m \in \mathbb{R}^{n\times n}$ contains, as its columns, the right-singular vectors. Note that we assume that $\boldsymbol{P}_m^T\boldsymbol{U}$ has full column rank in the following, which can be ensured by initializing GappyPOD+E with, e.g., the QDEIM interpolation points. If we add a sampling point, we obtain

$$\boldsymbol{P}_{m+1}^T\boldsymbol{U} = \begin{bmatrix}\boldsymbol{P}_m^T\boldsymbol{U}\\ \boldsymbol{u}_+\end{bmatrix} \in \mathbb{R}^{m+1\times n}\,,$$

[1]Note that we have changed the notation slightly here and added the subscript "m" to $\boldsymbol{P}$. This will help distinguish the sampling points matrix when new indices are added.

where $\boldsymbol{u}_+ \in \mathbb{R}^{1\times n}$ is the row of $\boldsymbol{U}$ that is selected by the new sampling point. Following the work by Zimmermann and Willcox [45], the change of the singular values of $\boldsymbol{P}_m^T\boldsymbol{U}$ to $\boldsymbol{P}_{m+1}^T\boldsymbol{U}$ can be understood via a symmetric rank-one update. We have

$$\boldsymbol{P}_{m+1}^T\boldsymbol{U} = \begin{bmatrix} \boldsymbol{V}_m & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_m \\ \boldsymbol{u}_+\boldsymbol{W}_m \end{bmatrix} \boldsymbol{W}_m^T .$$

The singular values of $\boldsymbol{P}_{m+1}^T\boldsymbol{U}$ are given by the square roots of the eigenvalues of $(\boldsymbol{P}_{m+1}^T\boldsymbol{U})^T(\boldsymbol{P}_{m+1}^T\boldsymbol{U})$, which we represent as

$$(\boldsymbol{P}_{m+1}^T\boldsymbol{U})^T(\boldsymbol{P}_{m+1}^T\boldsymbol{U}) = \boldsymbol{W}_m \left(\boldsymbol{\Sigma}_m^2 + \boldsymbol{W}_m^T\boldsymbol{u}_+^T\boldsymbol{u}_+\boldsymbol{W}_m\right) \boldsymbol{W}_m^T .$$

Define $\boldsymbol{\Lambda}_{m+1} = \boldsymbol{\Sigma}_m^2 + \boldsymbol{W}_m^T\boldsymbol{u}_+^T\boldsymbol{u}_+\boldsymbol{W}_m$. With $\bar{\boldsymbol{u}}_+ = \boldsymbol{W}_m^T\boldsymbol{u}_+^T$, we obtain $\boldsymbol{\Lambda}_{m+1} = \boldsymbol{\Sigma}_m^2 + \bar{\boldsymbol{u}}_+\bar{\boldsymbol{u}}_+^T$, which is a symmetric rank-one update to the diagonal matrix $\boldsymbol{\Sigma}_m^2$. The square roots of the eigenvalues of $\boldsymbol{\Lambda}_{m+1}$ are the singular values of $\boldsymbol{P}_{m+1}^T\boldsymbol{U}$.

Let $\lambda_1^{(m)}, \dots, \lambda_n^{(m)}$ be the eigenvalues of $\boldsymbol{\Sigma}_m^2$ and let $\lambda_1^{(m+1)}, \dots, \lambda_n^{(m+1)}$ be the eigenvalues of Λ_{m+1}, both listed in descending order. Our goal is now to select a row of $\boldsymbol{U}$ that maximizes the smallest eigenvalue $\lambda_n^{(m+1)}$. From Weyl's theorem [40, 41] we have that $\lambda_n^{(m+1)} \geq \lambda_n^{(m)}$, which shows that adding any sampling point will, at least, not increase $\|(\boldsymbol{P}_{m+1}^T\boldsymbol{U})^\dagger\|_2$ compared to $\|(\boldsymbol{P}_m^T\boldsymbol{U})^\dagger\|_2$.

5.2. Lower bounds for eigenvalues of updated matrices. We now use the results by Ipsen and Nadler in [26] to derive a heuristic strategy with the aim of selecting sampling points that lead to a fast increase of the smallest eigenvalue, i.e., to a fast decrease of $\|(\boldsymbol{P}_m^T\boldsymbol{U})^\dagger\|_2$.

Let $g = \lambda_{n-1}^{(m)} - \lambda_n^{(m)}$ be the eigengap. Note that we need $\lambda_{n-1}^{(m)} > \lambda_n^{(m)}$ in the following. Let $\boldsymbol{z}_n^{(m)} \in \mathbb{R}^n$ be the eigenvector of $\boldsymbol{\Sigma}_m^2$ corresponding to the smallest eigenvalue $\lambda_n^{(m)}$, with $\|\boldsymbol{z}_n^{(m)}\|_2 = 1$. In our case $\boldsymbol{z}_n^{(m)}$ is the nth canonical unit vector of dimension n because $\boldsymbol{\Sigma}_m^2$ is diagonal with diagonal elements ordered descending. Then, as shown in [26, Corollary 2.2],

$$\lambda_n^{(m+1)} \geq \lambda_n^{(m)} + \frac{1}{2}\left(g + \|\bar{\boldsymbol{u}}_+\|_2^2 - \sqrt{\left(g + \|\bar{\boldsymbol{u}}_+\|_2^2\right)^2 - 4g(\boldsymbol{z}_n^{(m)T}\bar{\boldsymbol{u}}_+)^2}\right) . \tag{37}$$

Observe that the bound (37) depends on the eigen*vector* corresponding to the smallest eigenvalue.

The bound (37) motivates us to add the rows of $\boldsymbol{U}$ that maximize

$$g + \|\bar{\boldsymbol{u}}_+\|_2^2 - \sqrt{\left(g + \|\bar{\boldsymbol{u}}_+\|_2^2\right)^2 - 4g(\boldsymbol{z}_n^{(m)T}\bar{\boldsymbol{u}}_+)^2} . \tag{38}$$

The criterion (38) is related to the criteria developed in [45]. While we build on the perturbation bounds introduced in [26], the authors of [45] directly derive criteria that take the eigenvector $\boldsymbol{z}_n$ corresponding to the smallest eigenvalue into account; see [45, p. A2834] and [45, Remark 2, item 3]. In fact, the work [45] goes a step further and also takes into account inner products with eigenvectors corresponding to larger eigenvalues. We do not consider these additional steps discussed in [45] in the following.

5.3. The GappyPOD+E algorithm. The GappyPOD+E sampling approach that we consider is summarized in Algorithm 1. It iteratively selects new sampling points that maximize (38) in a greedy fashion. In line 2 of Algorithm 1, the first n

points are selected with QDEIM; see Listing 1. Then, for each point $i = n+1, \ldots, m$, the SVD of $\boldsymbol{P}_i^T \boldsymbol{U}$ (which is `U(p, :)` in the notation used in Algorithm 1) is computed to obtain the right-singular vectors as columns of the matrix $\boldsymbol{W}_m$. The eigengap g is computed on line 6 in Algorithm 1. Then, $\bar{\boldsymbol{U}} = \boldsymbol{W}_i^T \boldsymbol{U}^T$ is obtained on line 7 in Algorithm 1. The bound (38) is then computed from $\bar{\boldsymbol{U}}$ for each column $\bar{\boldsymbol{u}}_+$ in lines 8–9 and sorted descending on line 10. On line 9, it is exploited that the eigenvector $\boldsymbol{z}_n^{(i)}$ is the nth canonical unit vector of dimension n and so $\boldsymbol{z}_n^{(i)T} \bar{\boldsymbol{u}}_+$ in (38) for column $\bar{\boldsymbol{u}}_+$ of $\bar{\boldsymbol{U}}$ is computed as $\boldsymbol{z}_n^{(i)T} \bar{\boldsymbol{u}}_+ = \boldsymbol{z}_n^{(i)} \boldsymbol{W}_i^T \boldsymbol{u}_+^T = (\boldsymbol{w}_i^{\text{end}})^T \boldsymbol{u}_+^T$, where $\boldsymbol{w}_i^{\text{end}}$ is the right-singular vector corresponding to the smallest singular value. This means that $(\boldsymbol{z}_n^{(i)})^T \bar{\boldsymbol{U}}$ is given by the last row of $\bar{\boldsymbol{U}}$ (denoted as `Ub(end, :)` in Algorithm 1). The point corresponding to the column of $\bar{\boldsymbol{U}}$ (row of $\boldsymbol{U}$) with the largest value (38) is added as a sampling point and the procedure is repeated. Each iteration in GappyPOD+E requires performing an SVD of a small matrix whose size grows with the reduced dimension n and the number of sampling points. Each SVD is in $\mathcal{O}(n^2 m)$ (for $m > n$). Thus, selecting m points with GappyPOD+E is in $\mathcal{O}(n^2 m^2)$. Note that the sampling point selection is performed during the construction of the reduced model in the offline phase.

The GappyPOD+E algorithm returns points that are not necessarily nested with respect to the dimension of the DEIM basis: Consider a basis matrix $\boldsymbol{U}_n$ with n columns and the corresponding set $\mathcal{P}_n$ of $m > n$ points selected by GappyPOD+E. Let now $\boldsymbol{U}_{n+1}$ be a basis matrix with $n+1$ columns where the first n columns coincide with the columns of $\boldsymbol{U}_n$ and let $\mathcal{P}_{n+1}$ be the set of at least m points selected with GappyPOD+E. Then, it is possible that $\mathcal{P}_n \not\subset \mathcal{P}_{n+1}$, which is in contrast to, e.g., the greedy EIM algorithm [6], for which $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ holds if n and $n+1$ points are selected, respectively. Nestedness of points is a desired property in situations where one wants to, for example, rapidly and adaptively select the number of DEIM basis vectors and sampling points without running the sampling algorithm from scratch. One such situation is in model reduction when the dimension of the DEIM space is selected during the online phase. One option to avoid running GappyPOD+E during the online phase in this situation is to precompute GappyPOD+E sampling points for a range of dimensions of the DEIM space and to store them during the offline phase, which might require a large amount of memory. In the online phase, the precomputed points can be quickly loaded depending on the basis dimension that is selected online, instead of running GappyPOD+E during the online phase.

6. Numerical results. This section compares the stability of DEIM and GappyPOD with randomized and deterministic sampling algorithms on numerical examples. Subsection 6.1 revisits the toy example from subsection 2.3 and demonstrates that GappyPOD provides stable approximations compared to DEIM. Subsection 6.2 approximates velocity fields from noisy measurements of single-injector combustion processes following the procedure introduced in [9, 42]. Subsection 6.3 demonstrates the effect of taking more sampling points than basis vectors on a diffusion-reaction problem, where GappyPOD provides stable approximations in contrast to DEIM.

6.1. Synthetic example. Let us revisit the synthetic example introduced in subsection 2.3. We use the same setup as before but now approximate the noisy function with GappyPOD that takes more sampling points than basis vectors. We compare our GappyPOD+E sampling strategy to three sampling strategies from the literature. First, there is GappyPOD+R, which takes the first n sampling points with QDEIM and the subsequent $m-n$ sampling points uniform randomly with replacement. Thus, the

Algorithm 1 Sampling points selection with GappyPOD+E (MATLAB notation).

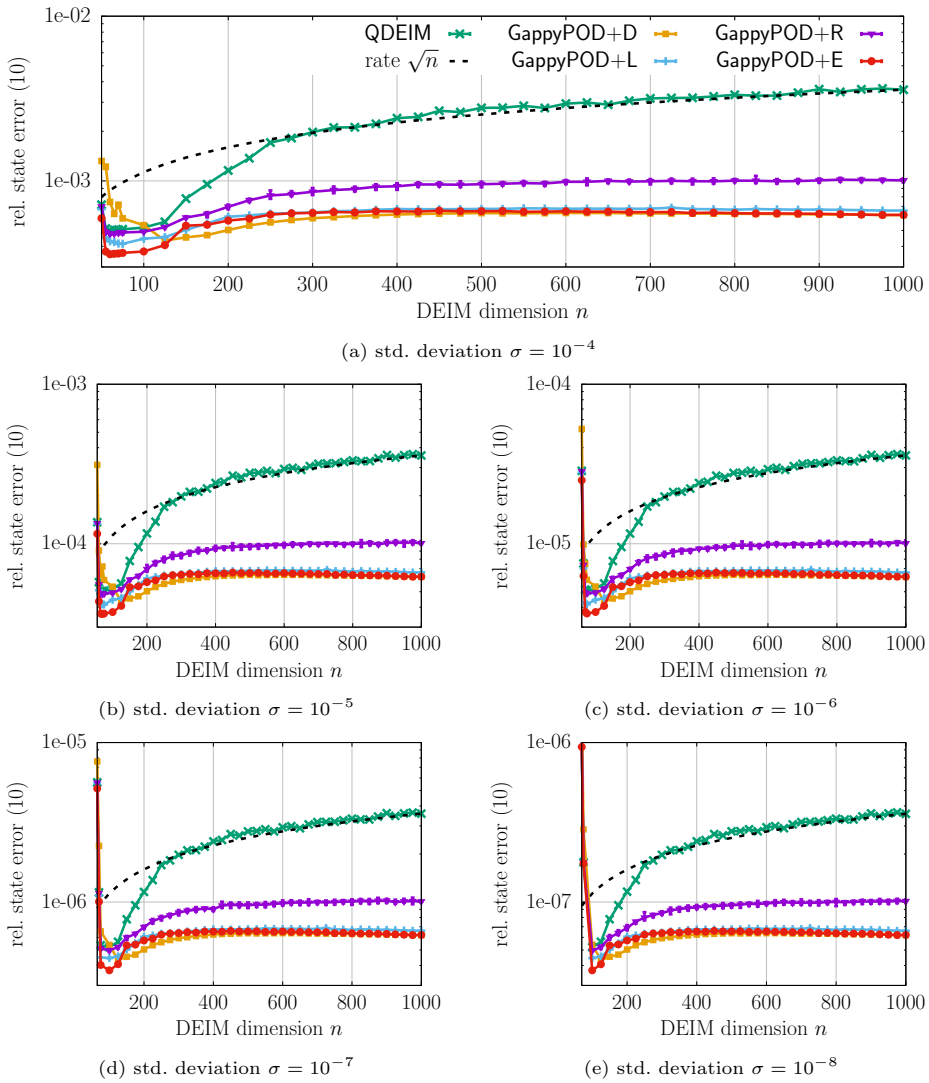
```
function [ p ] = gpode( U, m )
[~, ~, p] = qr(U', 'vector');
p = p(1:size(U, 2))';
for i=length(p)+1:m
    [~, S, W] = svd(U(p, :), 0);
    g = S(end-1, end-1).^2 - S(end, end)^2;
    Ub = W'*U';
    r = g + sum(Ub.^2, 1);
    r = r-sqrt((g+sum(Ub.^2,1)).^2-4*g*Ub(end, :).^2);
    [~, I] = sort(r, 'descend');
    e = 1;
    while any(I(e) == p)
        e = e + 1;
    end
    p(end + 1) = I(e);
end
end
```

"R" in GappyPOD+R stands for "random." Second, the strategy GappyPOD+L takes the first n points with QDEIM and the subsequent $m - n$ points based on leverages scores as described in, e.g., [3, section V.B]. Thus, the "L" in GappyPOD+L stands for "leverage scores." Third, with GappyPOD+D we denote the sampling strategy introduced in [13, Algorithm 4] that selects $m > n$ sampling points by extending the DEIM greedy algorithm [6, 14]. Thus, the "D" in GappyPOD+D stands for "DEIM greedy." The number of sampling points is set to $m = 2n$ in case of GappyPOD in the following. We perform 10 replicates of the experiments and compute the error as defined in (10). The results in Figure 2(a) indicate that GappyPOD with more sampling points than basis vectors avoids the unstable behavior obtained with DEIM, as suggested by our analysis presented in section 4. All sampling algorithms perform well in this example, with GappyPOD+E, GappyPOD+L, GappyPOD+D achieving the lowest errors. Similar results are obtained for $\sigma \in \{10^{-5}, \dots, 10^{-8}\}$ in this example as shown in Figure 2(b)–(e). Note that the error decays linearly with σ as long as the noise limits the overall accuracy rather than the projection error, as indicated by Theorem 4.6. The error bars in Figure 2 show the minimum and maximum of the error over the 10 replicates. Except for GappyPOD+R in Figure 2(a) near $n = 1000$, the error bars are barely visible in the plots, which indicates that even small perturbations due to noise lead to unstable behavior in QDEIM.

Figure 3(a) compares the orthogonal projection error of the noisy data,

$$\sum_{j=1}^{k} \frac{1}{2500} \sum_{i=1}^{2500} \frac{\|\boldsymbol{f}(\boldsymbol{x};\xi_i') - \boldsymbol{U}\boldsymbol{U}^T \boldsymbol{f}_{\boldsymbol{\epsilon}_j}(\boldsymbol{x};\xi_i')\|_2}{\|\boldsymbol{f}(\boldsymbol{x};\xi_i')\|_2}, \tag{39}$$

to error (10) of QDEIM and GappyPOD+R. Note that the projection error (39) grows with the dimension n since the Gaussian random noise vector can be better and better approximated in the subspace spanned by the columns of $\boldsymbol{U}$ as the dimension n is increased. Figure 3(b) shows that if the dimension $n = 500$ is fixed and the number



(a) std. deviation $\sigma = 10^{-4}$

(b) std. deviation $\sigma = 10^{-5}$

(c) std. deviation $\sigma = 10^{-6}$

(d) std. deviation $\sigma = 10^{-7}$

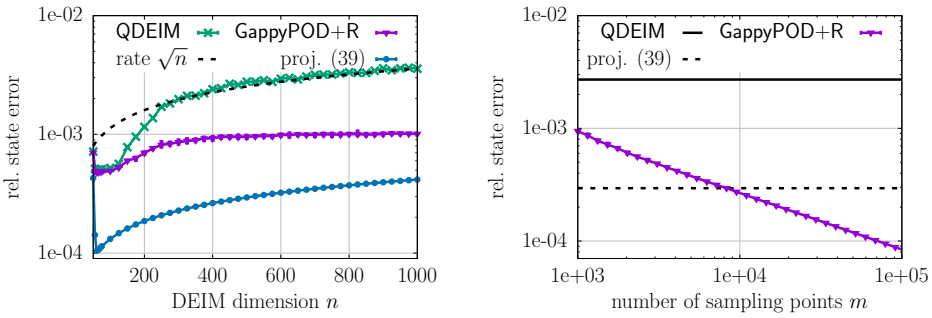
(e) std. deviation $\sigma = 10^{-8}$

FIG. 2. *Synthetic example:* GappyPOD *with more sampling points than basis vectors shows stable behavior and avoids the amplification of the error with a rate* $\sqrt{n}$ *as observed in the* QDEIM *approximation. All sampling strategies for* GappyPOD *give stable approximations in this example.*

of sampling points m is increased, then the error (10) of GappyPOD+R decays with a rate $\sqrt{1/m}$ to the projection error of noiseless data, i.e.,

$$\sum_{j=1}^{k} \frac{1}{2500} \sum_{i=1}^{2500} \frac{\|\boldsymbol{f}(\boldsymbol{x};\xi_i') - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f}(\boldsymbol{x};\xi_i')\|_2}{\|\boldsymbol{f}(\boldsymbol{x};\xi_i')\|_2}\,, \tag{40}$$

which is below machine precision in this example for $n = 500$. The results shown in Figure 3(b) are in alignment with Theorem 4.6, which states that the error of GappyPOD with uniform sampling with replacement converges with rate $\sqrt{1/m}$ to the projection error.



(a) sampling with ratio $m/n = 2$ (b) dimension $n = 500$ fixed

FIG. 3. *Plot* (a) *shows the projection error* (39) *of noisy data with standard deviation* $\sigma = 10^{-4}$. *The projection error* (39) *grows with the space dimension* n *because the Gaussian random noise vector can be more accurately approximated in the space as the dimension is increased. Plot* (b) *indicates that the error* (10) *of* GappyPOD+R *converges with a rate* $\sqrt{1/m}$ *to the projection error of noiseless data* (40) *for a fixed dimension; cf. Theorem* 4.6. *The projection error of noiseless data is below machine precision in this example.*

6.2. Velocity field approximations from noisy measurements of single-injector combustion process. We consider velocity field approximations from noisy measurements of the single-injector combustion process described in detail in [39]. The combustion model follows the implementation of the General Equation and Mesh Solver (GEMS) code [23, 25] developed by Purdue University. The domain of the setup of [39] is shown in Figure 4. Fuel and oxidizer are input with constant mass flow rates of $5.0\frac{\text{kg}}{s}$ and $0.37\frac{\text{kg}}{s}$, respectively. The fuel is composed of gaseous methane and the oxidizer is 42% gaseous O_2 and 58% gaseous H_2O. Details of the physics of the problem setup are described in [25].

To generate snapshot data, the GEMS code is used to simulate the system for 0.7ms with a time step size of $\delta t = 10^{-7}$. The simulation leads to 7000 snapshots $\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_{7000}) \in \mathbb{R}^{77046}$ at 7000 time points $t_1, \ldots, t_{7000}$. The snapshots are of length 77046 (there are 38523 spatial discretization points) and contain the velocity in the x and y directions. The basis matrix $\boldsymbol{U}$ is derived with POD from the snapshots, where every fourth snapshot is skipped and kept as a test snapshot. Thus, the basis matrix is constructed from 5250 snapshots that are the columns of $\boldsymbol{X}$ and the 1750 test snapshots are ignored during construction of the basis matrix and collected as columns in $\boldsymbol{X}^{(\text{test})}$. The test snapshots are polluted with zero-mean Gaussian noise with standard deviation $\sigma = 1.7$ and $\sigma = 3.4$, respectively, and collected in the

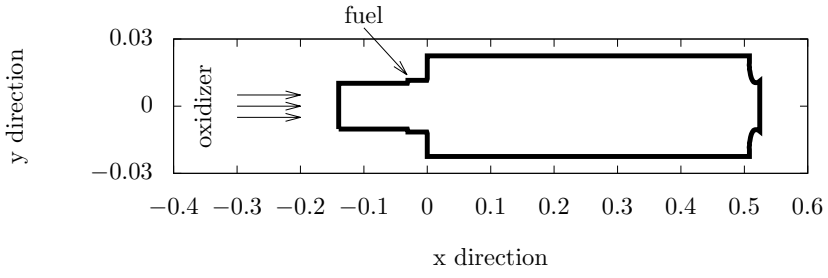
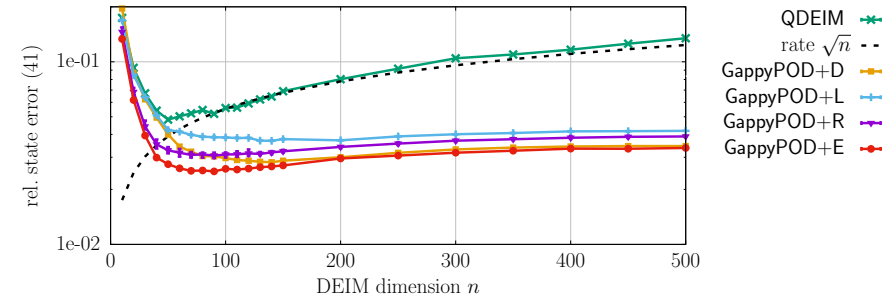


FIG. 4. *Combustion: Geometry of combustion chamber; see* [23, 25] *for details.*

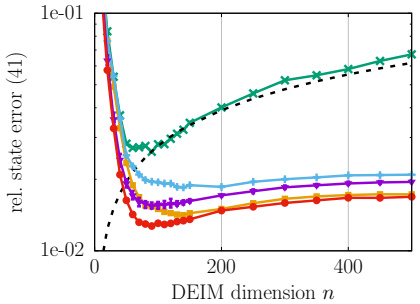
noisy test snapshot matrix $\boldsymbol{X}_{\boldsymbol{\epsilon}}^{(\text{test})}$. A standard deviation of $\sigma = 1.7$ corresponds to about 0.5% noise with respect to the mean of the snapshot matrix. Correspondingly, $\sigma = 3.4$ means that about 1% noise is added to the test snapshots. Figure 5(a) shows for $\sigma = 3.4$ the relative state error

$$\sum_{j=1}^{10} \frac{\left\| \boldsymbol{X}^{(\text{test})} - \boldsymbol{U}(\boldsymbol{P}^T\boldsymbol{U})^\dagger \boldsymbol{P}^T \boldsymbol{X}_{\boldsymbol{\epsilon}_j}^{(\text{test})} \right\|_F}{\|\boldsymbol{X}^{(\text{test})}\|_F} \tag{41}$$

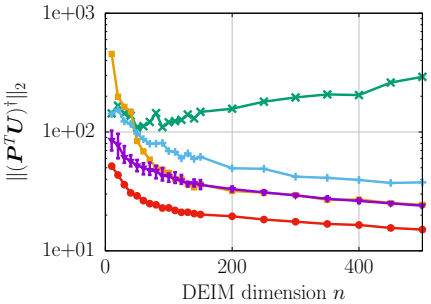
over $j = 1, \ldots, 10$ replicates of noise $\boldsymbol{\epsilon}_j$. The matrix $\boldsymbol{P}$ is derived from QDEIM with $m = n$ and from GappyPOD+D, GappyPOD+L, GappyPOD+R, GappyPOD+E, respectively, with $m = 2n$, i.e., twice as many sampling points as number of basis vectors. Figure 5(a) shows the growth of the error (41) for QDEIM, which uses the same number of sampling points as basis vectors. In contrast, taking more sampling points than basis vectors with GappyPOD yields a stable approximation. All sampling strategies help to reduce the error (41) significantly, where GappyPOD+E achieves the lowest error in this example. Error bars show the minimum and the maximum of the error over the replicates. The error bars are barely visible in Figure 5(a), which indicates that small perturbations can lead to unstable behavior in QDEIM and that GappyPOD with more sampling points than basis vectors robustly gives stable approximations. Similar results are obtained for $\sigma = 1.7$ in Figure 5(b). Figure 5(c)



(a) std. deviation $\sigma = 3.4$ (about 1% noise)



(b) std. deviation $\sigma = 1.7$ (about 0.5% noise)



(c) norm of sampling operator

FIG. 5. *Combustion: The plots in this figure show that approximating the velocity field of the single-injector combustion process considered in this example from noisy measurements suffers from the instability described in section 2.3 if QDEIM is used with the same number of sampling points as the dimension of the reduced space. In contrast, GappyPOD with various sampling strategies yields stable approximations, i.e., avoids the growth with rate $\sqrt{n}$ with the dimension n of the reduced space. GappyPOD+E achieves the lowest error in this example.*

shows that the norm of the sampling operator $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2$ is lowest for GappyPOD+E, which is in agreement with the results in Figures 5(a) and 5(b).

6.3. Diffusion-reaction problem with nonlinear reaction term. We now demonstrate the stability of GappyPOD and DEIM on a reduced model of a diffusion-reaction problem. The example demonstrates that instabilities in the DEIM approximations can lead to unstable reduced models, which can be avoided with GappyPOD if more sampling points than basis vectors are used.

6.3.1. Problem setup. Let $\Omega = (0,1)^2 \subset \mathbb{R}^2$ and $\mathcal{D} = [-\pi/2, \pi/2] \times [1,5]$ and consider the PDE

$$-\Delta u(\boldsymbol{\omega};\boldsymbol{\xi}) + f(u(\boldsymbol{\omega};\boldsymbol{\xi});\boldsymbol{\xi}) = 100\sin(2\pi\omega_1)\sin(2\pi\omega_2)\,, \qquad \boldsymbol{\omega} \in \Omega\,, \tag{42}$$

where $\boldsymbol{\omega} = [\omega_1, \omega_2]^T$ is the spatial coordinate, $u : \Omega \times \mathcal{D} \to \mathbb{R}$ is the solution function, and $f : \mathbb{R} \times \mathcal{D} \to \mathbb{R}$ is a nonlinear function

$$f(u;\boldsymbol{\xi}) = (0.1\sin(\xi_1) + 2)\exp(-2.7\xi_1^2)(\exp(\xi_2 u 1.8) - 1)$$

with parameter $\boldsymbol{\xi} = [\xi_1, \xi_2]^T \in \mathcal{D}$. The PDE (42) is closed with homogeneous Dirichlet boundary conditions. This example is a modification of the example considered in [22].

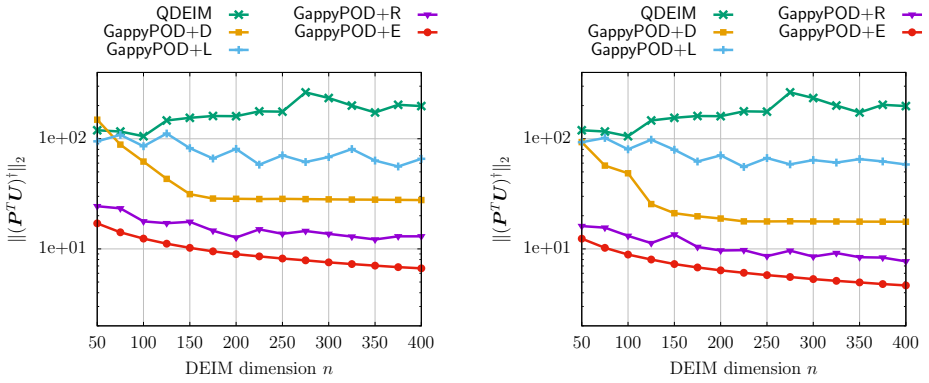
We discretize (42) with a second-order finite difference scheme on an equidistant mesh with mesh width $h = 1/255$ in Ω, which leads to the state dimension $N = 65536$. The system of nonlinear equations is solved with Newton's method. We derive a reduced model from 1600 snapshots corresponding to a 40×40 grid of parameter values in the domain $\mathcal{D}$. The grid is equidistant in the first direction and logarithmically equidistant in the second direction. The basis matrix $\boldsymbol{V}$ is constructed with POD. The POD dimension is chosen as $r = 50$. The nonlinear term is approximated with empirical interpolation, with more details to follow below. The reduced model is tested on parameters corresponding to the 9×9 grid in $\mathcal{D}$ that is linearly equidistant in the first direction and logarithmically equidistant in the second direction. The full-model states corresponding to the test parameters are collected in the test snapshot matrix $\boldsymbol{X}^{(\text{test})} \in \mathbb{R}^{N \times 81}$.

6.3.2. Results. We compare reduced models that differ in the way the nonlinear term is approximated. With "QDEIM" we denote the reduced models that approximate the nonlinear terms with QDEIM, which takes $m = n$ sampling points. Reduced models that approximate the nonlinear term with GappyPOD are denoted as "GappyPOD+D," "GappyPOD+R," "GappyPOD+L," and "GappyPOD+E," respectively, depending on which sampling strategy is used.

Figure 6 compares the norm of the sampling operators for $m = 4n$ and $m = 8n$ for dimensions $n \in \{50, \ldots, 400\}$. GappyPOD+E provides the sampling operator with the lowest norm in this example. We first run the reduced models for the test parameters without adding noise and collect the corresponding states as columns in the matrix $\tilde{\boldsymbol{X}}^{(\mathsf{S})} \in \mathbb{R}^{r \times 81}$ where S is either QDEIM, GappyPOD+D, GappyPOD+L, GappyPOD+R, or GappyPOD+E. The averaged relative state error

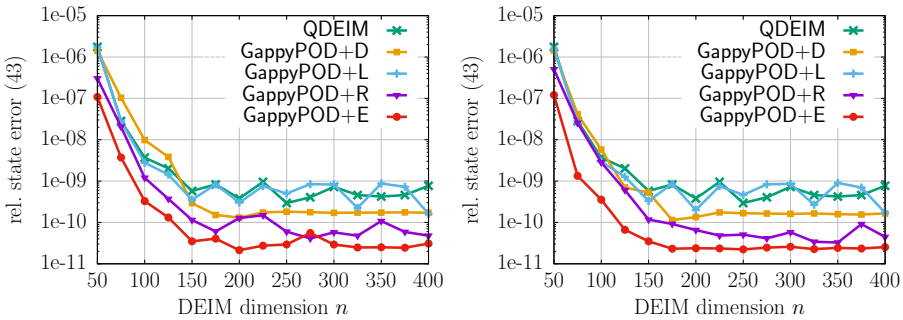
$$\frac{\|\boldsymbol{X} - \boldsymbol{V}\tilde{\boldsymbol{X}}^{(\mathsf{S})}\|_F}{\|\boldsymbol{X}\|_F} \tag{43}$$

is shown in Figure 7(a) for an oversampling factor $m/n = 4$ and in Figure 7(b) for $m/n = 8$ for S either QDEIM, GappyPOD+D, GappyPOD+L, GappyPOD+R, or GappyPOD+E. QDEIM as well as GappyPOD with all sampling strategies achieve



(a) $m = 4n$ (factor 4 oversampling) (b) $m = 8n$ (factor 8 oversampling)

FIG. 6. *Diffusion reaction example: The sampling operators derived with GappyPOD+E achieves the lowest norm* $\|(\boldsymbol{P}^T\boldsymbol{U})^\dagger\|_2$ *in this example.*



(a) $m = 4n$ (factor 4 oversampling) (b) $m = 8n$ (factor 8 oversampling)

FIG. 7. *Diffusion reaction example: Without noise, QDEIM and GappyPOD show stable behavior in this example. Note that the dimension* r *of the POD space is fixed and therefore the curves level off even if the dimension* n *of the reduced space spanned by the columns of* $\boldsymbol{U}$ *is increased.*

stable approximations in the sense described in section 2.3, i.e., the error (43) does not grow with the dimension n of the reduced space.

We now run the reduced models for the test parameters and perturb the nonlinear function evaluations f with zero-mean Gaussian noise and standard deviation $\sigma > 0$. We repeat this process $k = 10$ times and collect the states of a reduced model corresponding to the test parameters as columns in $\tilde{\boldsymbol{X}}_i^{(\mathsf{S})} \in \mathbb{R}^{r\times 81}$ for $i = 1, \ldots, k$. Then, the averaged relative state error

$$\sum_{i=1}^{k} \frac{\|\boldsymbol{X} - \boldsymbol{V}\tilde{\boldsymbol{X}}_i^{(\mathsf{S})}\|_F}{\|\boldsymbol{X}\|_F} \tag{44}$$

is reported in the following for each reduced model. Figure 8 compares the error (44) for reduced models based on QDEIM, GappyPOD+D, GappyPOD+L, GappyPOD+R, and GappyPOD+E. The standard deviation of the noise is $\sigma = 10^{-2}$ and the oversampling factor is 4, i.e., $m = 4n$. The growth of the error (44) with rate $\sqrt{n}$ can be observed for QDEIM in Figure 8. Similarly, the reduced models based on GappyPOD+D seem unstable because the corresponding errors grow with a rate of

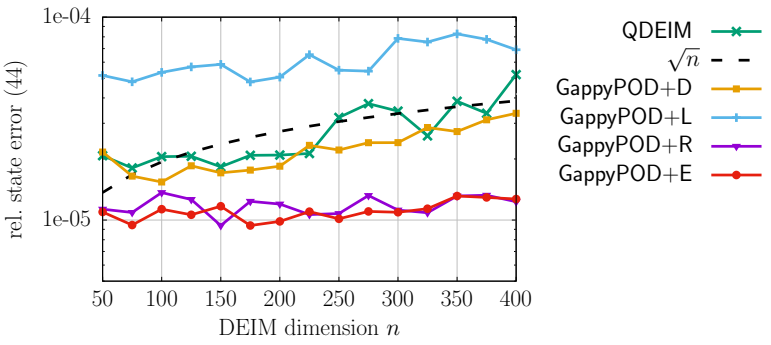
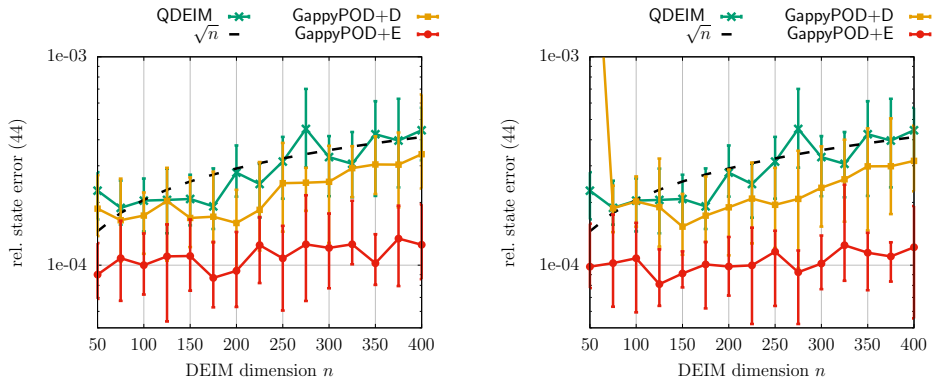


FIG. 8. *Diffusion reaction example with noise: Approximating the nonlinear terms in this example with QDEIM leads to unstable behavior, which is indicated in this plot with a growth with rate* $\sqrt{n}$. *GappyPOD with more sampling points than basis vectors leads to stable reduced models with sampling strategies GappyPOD+E and GappyPOD+R in this example. Standard deviation of noise is* $\sigma = 10^{-2}$ *and oversampling factor is* $m/n = 4$.

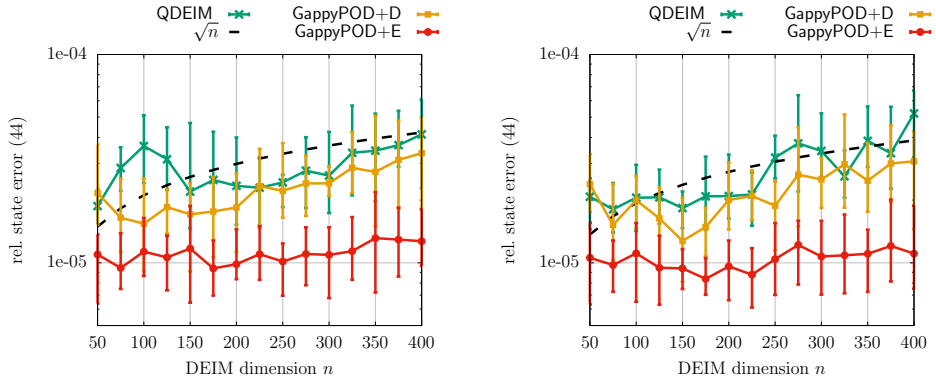
$\sqrt{n}$ too. In contrast, GappyPOD+E and GappyPOD+R give stable reduced models, where the error does not increase with the dimension n of the reduced space spanned by the columns of $\boldsymbol{U}$. The curves plotted in Figure 8 are shown in Figure 11 in the appendix with error bars that indicate the minimum and maximum errors over the $k = 10$ replicates. The sampling points selected with GappyPOD+L lead to models with poor performance in this example even though the growth of the error with rate $\sqrt{n}$ cannot be observed in the plot in Figure 8. However, the error bars shown in Figure 11 for GappyPOD+L are larger than for the other sampling algorithms, which indicates that there is strong variability in the approximation error achieved with GappyPOD+L in this example. The strong variability with respect to accuracy of the selected points might hide the growth of the error. Figure 9 compares GappyPOD+E with QDEIM and GappyPOD+D for oversampling factors $m/n = 4$ and $m/n = 8$ and standard deviations $\sigma \in \{10^{-2}, 10^{-3}, 10^{-4}\}$. The error bars show the minimum and maximum errors over the $k = 10$ replicates. In all cases, GappyPOD+E leads to a stable reduced model in the sense that the error does not grow with the dimension n of the DEIM space, whereas QDEIM and GappyPOD+D show unstable behavior and a growth of the error with rate $\sqrt{n}$.

Consider now Figure 10 that shows results for POD dimension $r = 9$, which is lower than dimension $r = 50$ used previously. The POD space of dimension $r = 9$ preserves about 99.9% of the energy, a typical threshold used in model reduction; cf. [7, section 3.1.1]. Note that the energy is $\sum_{i=1}^{r} \zeta_i^2 / \sum_{i=1}^{N} \zeta_i^2$, where $\zeta_1, \ldots, \zeta_N$ are the singular values of the snapshot matrix in descending order. The standard deviation of the noise is set to $\sigma = 10^{-1}$ and $\sigma = 10^{-2}$, respectively. The mean over 10 runs is shown in Figure 10. Similar behavior in terms of error as for higher POD dimensions is observed. Plot (c) in Figure 10 shows a detail of (a) and indicates that the approximations based on GappyPOD+R and GappyPOD+E have less oscillatory error than approximations obtained with QDEIM, GappyPOD+L, and GappyPOD+D for DEIM dimensions n between 5 and 50 in this example.

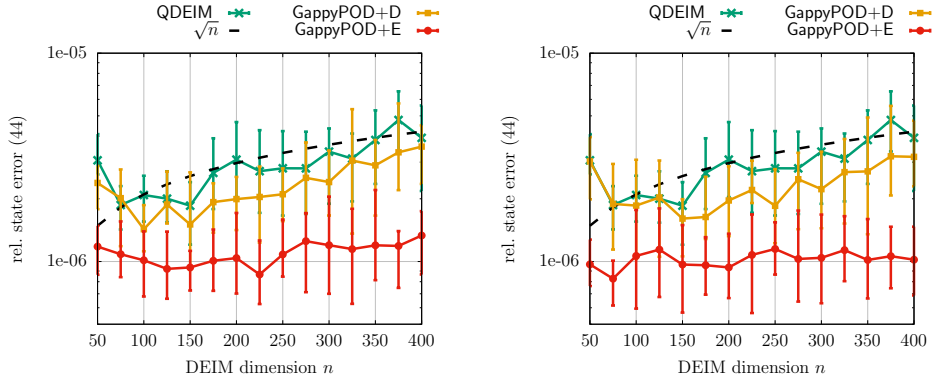
Remark 6.1. We comment on the problem setup. In this example, we considered nonlinear function evaluations that are perturbed with noise. We might encounter such a situation if, for example, parameters of the nonlinear function first need to be estimated from data via a Bayesian approach that introduces noise into the function



(a) std. deviation $\sigma = 10^{-2}$, oversampling $m = 4n$ (b) std. deviation $\sigma = 10^{-2}$, oversampling $m = 8n$



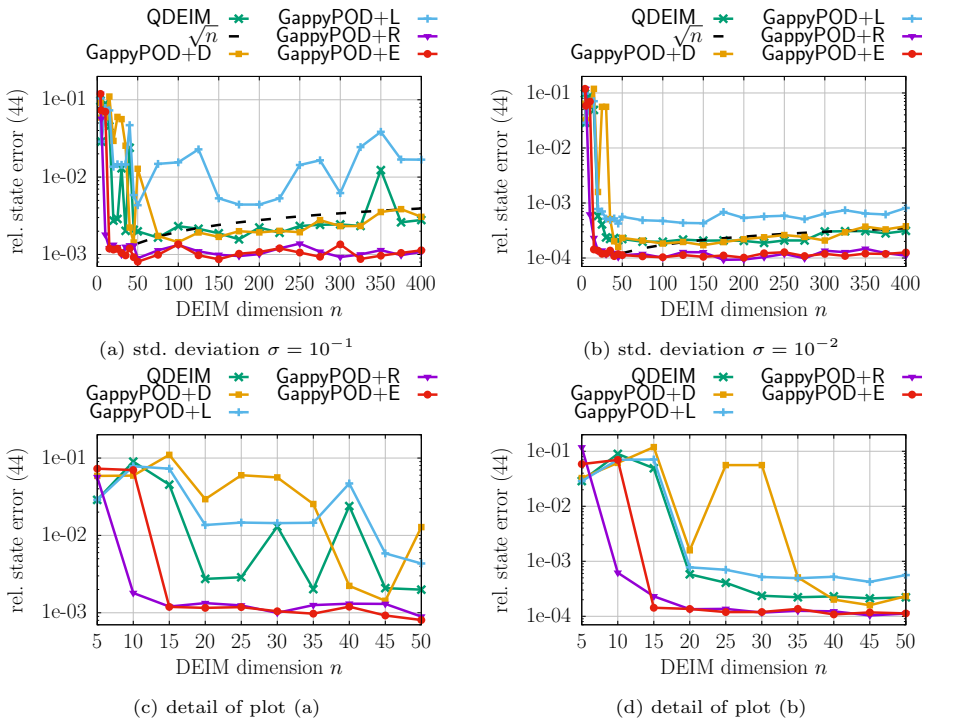
(c) std. deviation $\sigma = 10^{-3}$, oversampling $m = 4n$ (d) std. deviation $\sigma = 10^{-3}$, oversampling $m = 8n$



(e) std. deviation $\sigma = 10^{-4}$, oversampling $m = 4n$ (f) std. deviation $\sigma = 10^{-4}$, oversampling $m = 8n$

FIG. 9. *Diffusion reaction example: Taking more sampling points than basis vectors with GappyPOD+E leads to stable reduced models in this example. In contrast, reduced models based on QDEIM and GappyPOD+D exhibit instabilities in the sense of section* 2.3*, which is indicated by the growth of the error with the rate* $\sqrt{n}$*. Error bars show the minimum and maximum errors over* 10 *replicates.*

evaluations used in the reduced model. Our analysis does not cover deterministic approximation errors stemming from, e.g., relaxed tolerances of iterative solvers, and thus it remains future work to show if our analysis applies to such general types of noise as well.



(a) std. deviation $\sigma = 10^{-1}$

(b) std. deviation $\sigma = 10^{-2}$

(c) detail of plot (a)

(d) detail of plot (b)

FIG. 10. *Diffusion reaction example: The plots show that GappyPOD+E and GappyPOD+R achieve smaller errors than QDEIM also for a lower POD dimension $n = 9$ in this example. Notice in the detail of plot* (a) *shown in* (c) *that GappyPOD+R and GappyPOD+E show less oscillatory error behavior than QDEIM, GappyPOD+L, and GappyPOD+D for DEIM dimensions r between* 5 *and* 50.

7. Conclusions. Empirical interpolation is widely used for approximating nonlinear terms in reduced models and for recovering state fields from few spatial measurements; however, stability issues have been observed in presence of noise and other perturbations. Our probabilistic analysis shows that the particular instability that arises due to perturbations such as noise can be provably avoided by employing GappyPOD and taking more sampling points than dimensions of the reduced space. Numerical results demonstrated that instabilities in DEIM can lead to a loss of accuracy in the reduced model outputs and that GappyPOD with randomized and deterministic sampling strategies gives stabler approximations. The proposed deterministic sampling strategy aims to keep the number of required samples low and so directly affects the practical performance of the approximations.

Appendix A. Additional listing and figure.

LISTING 1
Selecting interpolation points with QDEIM [17] *(MATLAB code).*

```
function [ q ] = qdeim( U, m )
n = size(U, 2);
[~, ~, q] = qr(U', 'vector');
q = q(1:n)';
end
```

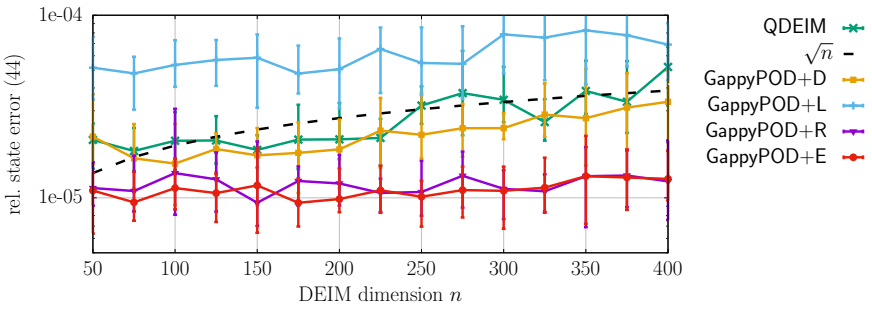


FIG. 11. *Diffusion reaction example: Approximating the nonlinear terms in this example with QDEIM leads to unstable behavior, which is indicated in this plot with a growth with rate* $\sqrt{n}$. *GappyPOD with more sampling points than basis vectors leads to stable reduced models with sampling strategies GappyPOD+E and GappyPOD+R in this example. Standard deviation of noise is* $\sigma = 10^{-2}$ *and oversampling factor is* $m/n = 4$.

Acknowledgments. The authors thank Karthik Duraisamy (University of Michigan), Cheng Huang (University of Michigan), and David Xu (University of Michigan) for providing the snapshots corresponding to the single-injector combustion process discussed in subsection 6.2.

REFERENCES

[1] J.-P. ARGAUD, B. BOURIQUET, F. DE CASO, H. GONG, Y. MADAY, AND O. MULA, *Sensor placement in nuclear reactors based on the generalized empirical interpolation method*, J. Comput. Phys., 363 (2018), pp. 354–370.

[2] J. P. ARGAUD, B. BOURIQUET, H. GONG, Y. MADAY, AND O. MULA, *Stabilization of (G)EIM in Presence of Measurement Noise: Application to Nuclear Reactor Physics*, in Spectral and High Order Methods for Partial Differential Equations, ICOSAHOM 2016, M. L. Bittencourt, N. A. Dumont, and J. S. Hesthaven, eds., Springer, Cham, 2017, pp. 133–145.

[3] P. ASTRID, S. WEILAND, K. WILLCOX, AND T. BACKX, *Missing point estimation in models described by proper orthogonal decomposition*, in Proceedings of the 43rd IEEE Conference on Decision and Control, Vol. 2, 2004, pp. 1767–1772.

[4] P. ASTRID, S. WEILAND, K. WILLCOX, AND T. BACKX, *Missing point estimation in models described by proper orthogonal decomposition*, IEEE Trans. Automat. Control, 53 (2008), pp. 2237–2251.

[5] L. BALZANO, B. RECHT, AND R. NOWAK, *High-dimensional matched subspace detection when data are missing*, in Proceedings of the IEEE International Symposium on Information Theory, 2010, pp. 1638–1642.

[6] M. BARRAULT, Y. MADAY, N. NGUYEN, AND A. PATERA, *An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations*, C. R. Math. 339 (2004), pp. 667–672.

[7] P. BENNER, S. GUGERCIN, AND K. WILLCOX, *A survey of projection-based model reduction methods for parametric dynamical systems*, SIAM Rev., 57 (2015), pp. 483–531.

[8] P. BINEV, A. COHEN, O. MULA, AND J. NICHOLS, *Greedy algorithms for optimal measurements selection in state estimation using reduced models*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 1101–1126.

[9] T. BUI-THANH, M. DAMODARAN, AND K. WILLCOX, *Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition*, AIAA J., 42 (2004), pp. 1505–1516.

[10] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), 717.

[11] K. CARLBERG, *Model Reduction of Nonlinear Mechanical Systems via Optimal Projection and Tensor Approximation*, Ph.D. thesis, Stanford University, 2011.

[12] K. CARLBERG, C. BOU-MOSLEH, AND C. FARHAT, *Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations*, Internat. J. Numer. Methods Engrg., 86 (2011), pp. 155–181.

[13] K. Carlberg, C. Farhat, J. Cortial, and D. Amsallem, *The GNAT method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows*, J. Comput. Phys., 242 (2013), pp. 623–647.
[14] S. Chaturantabut and D. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM J. Sci. Comput., 32 (2010), pp. 2737–2764.
[15] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone, *Discrete least squares polynomial approximation with random evaluations—application to parametric and stochastic elliptic PDEs*, ESAIM Math. Model. Numer. Anal., 49 (2015), pp. 815–837.
[16] A. Cohen, M. A. Davenport, and D. Leviatan, *On the stability and accuracy of least squares approximations*, Found. Comput. Math., 13 (2013), pp. 819–834.
[17] Z. Drmač and S. Gugercin, *A new selection operator for the Discrete Empirical Interpolation Method—improved a priori error bound and extensions*, SIAM J. Sci. Comput., 38 (2016), pp. A631–A648.
[18] Z. Drmač and A. Saibaba, *The discrete empirical interpolation method: Canonical structure and formulation in weighted inner product spaces*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1152–1180.
[19] J. L. Eftang and B. Stamm, *Parameter multi-domain 'hp' empirical interpolation*, Internat. J. Numer. Methods Engrg., 90 (2012), pp. 412–428.
[20] R. Everson and L. Sirovich, *The Karhunen-Loeve procedure for gappy data*, J. Opt. Soc. Amer., 12 (1995), pp. 1657–1664.
[21] F. Ghavamian, P. Tiso, and A. Simone, *POD-DEIM model order reduction for strain-softening viscoplasticity*, Comput. Methods Appl. Mech. Engrg., 317 (2017), pp. 458–479.
[22] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, ESAIM Math. Model. Numer. Anal., 41 (2007), pp. 575–605.
[23] M. E. Harvazinski, C. Huang, V. Sankaran, T. W. Feldman, W. E. Anderson, C. L. Merkle, and D. G. Talley, *Coupling between hydrodynamics, acoustics, and heat release in a self-excited unstable combustor*, Phys. Fluids, 27 (2015), 045102.
[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, New York, 2012.
[25] C. Huang, K. Duraisamy, and C. Merkle, *Investigations and improvement of robustness of reduced-order models of reacting flow*, in Proceedings of the AIAA Scitech 2019 Forum, 2019.
[26] I. C. F. Ipsen and B. Nadler, *Refined perturbation bounds for eigenvalues of Hermitian and Non-Hermitian matrices*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 40–53.
[27] Y. Maday and O. Mula, *A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation*, in Analysis and Numerics of Partial Differential Equations, F. Brezzi, P. Colli Franzone, U. Gianazza, and G. Gilardi, eds., Springer, New York, 2013, pp. 221–235.
[28] Y. Maday, O. Mula, A. Patera, and M. Yano, *The generalized empirical interpolation method: Stability theory on Hilbert spaces with an application to the Stokes equation*, Comput. Methods Appl. Mech. Engrg., 287 (2015), pp. 310–334.
[29] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, *Data-Driven Sparse Sensor Placement for Reconstruction*, arXiv:1701.07569, 2017.
[30] K. Manohar, S. L. Brunton, and J. N. Kutz, *Environment identification in flight using sparse approximation of wing strain*, J. Fluids Structures, 70 (2017), pp. 162–180.
[31] G. Migliorati, F. Nobile, and R. Tempone, *Convergence estimates in probability and in expectation for discrete least squares with noisy evaluations at random points*, J. Multivariate Anal., 142 (2015), pp. 167–182.
[32] B. Peherstorfer, *Model Reduction for Transport-dominated Problems via Online Adaptive Bases and Adaptive Sampling*, arXiv:1812.02094, 2018.
[33] B. Peherstorfer, D. Butnaru, K. Willcox, and H. Bungartz, *Localized discrete empirical interpolation method*, SIAM J. Sci. Comput., 36 (2014), pp. A168–A192.
[34] B. Peherstorfer and K. Willcox, *Online adaptive model reduction for nonlinear systems via low-rank updates*, SIAM J. Sci. Comput., 37 (2015), pp. A2123–A2150.
[35] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Rev., 60 (2018), pp. 550–591.
[36] S. B. Pope, *Simple models of turbulent flows*, Phys. Fluids, 23 (2011), 011301.
[37] G. Rozza, D. Huynh, and A. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations*, Arch. Comput. Methods Eng., 15 (2007), pp. 1–47.

[38] S. Sargsyan, S. L. Brunton, and J. N. Kutz, *Nonlinear model reduction for dynamical systems using sparse sensor locations from learned libraries*, Phys. Rev. E, 92 (2015), p. 033304.

[39] R. Swischuk, B. Kramer, C. Huang, and K. Willcox, *Learning physics-based reduced-order models for a single-injector combustion process*, AIAA J., to appear.

[40] H. Weyl, *Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)*, Math. Ann., 71 (1912), pp. 441–479.

[41] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, New York, 1988.

[42] K. Willcox, *Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition*, Comput. & Fluids, 35 (2006), pp. 208–226.

[43] Y. B. Zhou, *Model Reduction for Nonlinear Dynamical Systems with Parametric Uncertainties*, Ph.D. thesis, Massachusetts Institute of Technology, 2012.

[44] R. Zimmermann, B. Peherstorfer, and K. Willcox, *Geometric subspace updates with applications to online adaptive nonlinear model reduction*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 234–261.

[45] R. Zimmermann and K. Willcox, *An accelerated greedy missing point estimation procedure*, SIAM J. Sci. Comput., 38 (2016), pp. A2827–A2850.