

Article

A Stochastic Discrete Empirical Interpolation Approach for Parameterized Systems

Daheng Cai ¹ , Chengbin Yao ² and Qifeng Liao ^{1,*} 
¹ School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China; caidh@shanghaitech.edu.cn

² College of Natural Resources and Environment, Northwest A&F University, Xianyang 712100, China; chambiony@nwafu.edu.cn

* Correspondence: liaofq@shanghaitech.edu.cn

Abstract: As efficient separation of variables plays a central role in model reduction for nonlinear and nonaffine parameterized systems, we propose a stochastic discrete empirical interpolation method (SDEIM) for this purpose. In our SDEIM, candidate basis functions are generated through a random sampling procedure, and the dimension of the approximation space is systematically determined by a probability threshold. This random sampling procedure avoids large candidate sample sets for high-dimensional parameters, and the probability based stopping criterion can efficiently control the dimension of the approximation space. Numerical experiments are conducted to demonstrate the computational efficiency of SDEIM, which include separation of variables for general nonlinear functions, e.g., exponential functions of the Karhu nen–Loève (KL) expansion, and constructing reduced order models for FitzHugh–Nagumo equations, where symmetry among limit cycles is well captured by SDEIM.

Keywords: approximation theory; discrete empirical interpolation method; empirical interpolation method; variable separation


Citation: Cai, D.; Yao, C.; Liao, Q.

A Stochastic Discrete Empirical Interpolation Approach for Parameterized Systems. *Symmetry* **2022**, *14*, 556. <https://doi.org/10.3390/sym14030556>

Academic Editor: Alina Alb Lupas

Received: 15 February 2022

Accepted: 5 March 2022

Published: 10 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When conducting model reduction for nonlinear and nonaffine parameterized systems [1], separation of variables is an important step. During the last few decades, many strategies have been developed to achieve this goal, e.g., empirical interpolation methods (EIM) [2–5], discrete empirical interpolation methods (DEIM) [6,7], and variable-separation (VS) methods [8]. To result in an accurate linear representation, these methods typically need a fine reduced basis approximation space. To construct the corresponding basis functions, repeated evaluations of expensive parameterized systems are required. The accuracy of the EIM/DEIM approximations depends on candidate parameter samples. Properly choosing the samples is crucial and is especially challenging when the parameter space is high dimensional.

For reduced basis approximations [9–14], the work in [15] shows that candidate sample sets for reduced bases can be chosen as a random set of a specified order and the resulting approximation satisfies given accuracies with a probability threshold. Rather than reduced basis approximations, we in this paper focus on separation of variables for nonlinear and nonaffine systems, and propose a stochastic discrete empirical interpolation method (SDEIM). In our SDEIM, the interpolation is processed through two steps: the first is to randomly select sample points to construct an approximation space for empirical interpolation, and the second is to evaluate if the approximation accuracy meets the probability threshold on additional samples. These two steps are repeated until the approximation space satisfies the given accuracy and probability requirements. The probability correctness of the approximation space is reassessed by SDEIM each time the approximation space is

updated. We note that SDEIM does not loop over a fine candidate sample set, which significantly reduces the computational cost for empirical interpolation for high-dimensional systems. To demonstrate the efficiency of SDEIM, we utilize it to approximate test functions including the exponential functions of the Karhu nen–Loève (KL) expansion, which is widely used to parameterize random fields. Finally, we use SDEIM to solve ordinary differential equations (ODEs) arising from the FitzHugh–Nagumo (F–N) system [6,16], and all these results show that SDEIM is efficient. It is worth noting that many natural phenomena can have symmetric properties, and symmetry of limit cycles of the F–N system is well captured by SDEIM.

An outline of the paper is as follows. We first present our problem setting, and then review the fine collateral reduced-basis approximation space and EIM in Section 2. After a short introduction of the discrete form, we present SDEIM and analyze its performance in Section 3. Numerical results are discussed in Section 4. Finally, we conclude the paper in Section 5.

2. Problem Formulation

Nonlinear terms and functions in complex systems cause significant difficulties for efficient model reduction, and separation of variables is typically required. In this section, we refer to nonlinear functions under consideration as the target functions. Let $\Omega \subset \mathbb{R}^d$ (for $d = 1, 2$ or 3) be a bounded, connected physical domain and Γ be a high dimensional parameter space. The target function with a general form is written as

$$\xi \mapsto f(x, \xi), x \in \Omega, \xi \in \Gamma.$$

The target function $f(x, \xi)$ is assumed to be nonlinear. In the following, we review EIM [2], which is to approximate $f(x, \xi)$ with a separate representation for x and ξ .

2.1. The Linear Approximation Space

Before introducing EIM, we first introduce its fine collateral reduced-basis approximation space following [2,6,17,18]. Without loss of generality, we assume that the target function $f(x, \xi)$ is uniquely defined in some Hilbert space \mathcal{H} for every $\xi \in \Gamma$. An inner product and a L_2 norm of \mathcal{H} are denoted as $\langle f, g \rangle := \int_{\Omega} f(x, \xi)g(x, \xi) dx$ and $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$. We define the target manifold \mathcal{M} as

$$\mathcal{M} := \{f(x, \xi) \mid \xi \in \Gamma\}.$$

when the target manifold \mathcal{M} has a low dimension, a low dimensional linear space can approximate it well. The infimum of the supremum distance between the n -dimensional linear approximation space and the target manifold \mathcal{M} is called the Kolmogorov n -width [3,19], which is defined as

$$\kappa_n = \kappa_n(\mathcal{M}) := \inf_{\dim(\mathcal{V})=n} \sup_{\xi \in \Gamma} \|f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi)\|. \quad (1)$$

Here, $\mathcal{P}_{\mathcal{V}}$ denotes the orthogonal projection onto \mathcal{V} , where \mathcal{V} is an arbitrary linear space, i.e.,

$$\langle \mathcal{P}_{\mathcal{V}}f(x, \xi), f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi) \rangle = 0. \quad (2)$$

Moreover, from (2) we can get the property

$$\begin{aligned} \|\mathcal{P}_{\mathcal{V}}f(x, \xi)\|^2 &= \langle \mathcal{P}_{\mathcal{V}}f(x, \xi), \mathcal{P}_{\mathcal{V}}f(x, \xi) \rangle \\ &= \langle f(x, \xi), f(x, \xi) \rangle - \langle f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi), f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi) \rangle \\ &= \|f(x, \xi)\|^2 - \|f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi)\|^2 \\ &\leq \|f(x, \xi)\|^2. \end{aligned} \quad (3)$$

Since the orthogonal projection is linear, (3) means that the error is reduced after the orthogonal projection. However, the space satisfying (1) is nontrivial to reach. The work in [20] has proposed a specific greedy procedure for constructing this n dimensional linear approximation space which optimizes

$$\tilde{\kappa}_n := \inf_{\dim(\mathcal{V})=n} \sup_{\xi \in \Gamma_{\text{train}}} \|f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi)\| \quad (4)$$

instead of κ_n in a sufficiently fine candidate sample set $\Gamma_{\text{train}} \subset \Gamma$. Since the candidate set Γ_{train} contains a finite number of samples, optimizing $\tilde{\kappa}_n$ is equivalent to search $\xi^{(k)}$ such that

$$\xi^{(k)} = \operatorname{argmax}_{\xi \in \Gamma_{\text{train}}} \|f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi)\|.$$

Then, we update the linear approximation space \mathcal{V} for this target function $f(x, \xi^{(k)})$. A greedy procedure which constructs a linear approximation space \mathcal{V} satisfying $\tilde{\kappa}_n \leq \frac{\text{tol}}{3}$, where tol is a given tolerance, can be stated as follows. First this procedure initializes the linear approximation space $\mathcal{V}_0 := \text{span}(\{0\})$. Then, it selects a new parameter value by

$$\xi^{(k+1)} := \operatorname{argmax}_{\xi \in \Gamma_{\text{train}}} \|f(x, \xi) - \mathcal{P}_{\mathcal{V}_k}f(x, \xi)\|.$$

If the error

$$\tilde{\eta}(\xi^{(k+1)}) = \|f(x, \xi^{(k+1)}) - \mathcal{P}_{\mathcal{V}_k}f(x, \xi^{(k+1)})\|$$

is greater than $\frac{\text{tol}}{3}$, we update the linear approximation space $\mathcal{V}_{k+1} = \text{span}(\mathcal{V}_k \cup f(x, \xi^{(k+1)}))$. At the end of this procedure, the linear approximation space \mathcal{V} is set to $\mathcal{V} = \mathcal{V}_n$ and \mathcal{V} satisfies $\tilde{\eta}(\xi) \leq \frac{\text{tol}}{3}$ for all $\xi \in \Gamma_{\text{train}}$. Moreover, that the candidate sample set Γ_{train} is sufficiently fine means that the discrete target manifold

$$\widetilde{\mathcal{M}} = \{f(x, \xi) | \xi \in \Gamma_{\text{train}}\}$$

is a $\frac{\text{tol}}{3}$ -approximation net for target manifold \mathcal{M} , i.e.,

$$d(\mathcal{M}, \widetilde{\mathcal{M}}) := \sup_{\xi \in \Gamma} \inf_{\tilde{\xi} \in \Gamma_{\text{train}}} \|f(x, \xi) - f(x, \tilde{\xi})\| \leq \frac{\text{tol}}{3}. \quad (5)$$

Then for any $\xi \in \Gamma$, (3) and (5) give

$$\begin{aligned} \|f(x, \xi) - \mathcal{P}_{\mathcal{V}}f(x, \xi)\| &\leq \inf_{\tilde{\xi} \in \Gamma_{\text{train}}} \left(\|f(x, \xi) - f(x, \tilde{\xi})\| + \|f(x, \tilde{\xi}) - \mathcal{P}_{\mathcal{V}}f(x, \tilde{\xi})\| \right. \\ &\quad \left. + \|\mathcal{P}_{\mathcal{V}}f(x, \tilde{\xi}) - \mathcal{P}_{\mathcal{V}}f(x, \xi)\| \right) \\ &\leq 2 \inf_{\tilde{\xi} \in \Gamma_{\text{train}}} \|f(x, \xi) - f(x, \tilde{\xi})\| + \frac{\text{tol}}{3} \\ &\leq \frac{2\text{tol}}{3} + \frac{\text{tol}}{3} = \text{tol}. \end{aligned} \quad (6)$$

The inequality (6) means that the standard greedy procedure with a fine candidate sample set Γ_{train} can generate a linear approximation space which approximates the target manifold \mathcal{M} with any tolerance tol .

2.2. Empirical Interpolation Method (EIM)

Section 2.1 explains that, when the candidate sample set Γ_{train} is fine enough, the greedy procedure can produce a linear approximation space \mathcal{V} which is a *tol*-approximation space for the original target manifold \mathcal{M} , i.e.,

$$\sup_{f(x,\xi) \in \mathcal{M}} \|f(x,\xi) - \mathcal{P}_{\mathcal{V}}f(x,\xi)\| \leq \text{tol}. \quad (7)$$

For this linear approximation space \mathcal{V} , if we denote a set of orthonormal basis functions of \mathcal{V} as $\{\phi_1(x), \dots, \phi_n(x)\}$, the orthogonal projection $\mathcal{P}_{\mathcal{V}}$ can be expressed as

$$\mathcal{P}_{\mathcal{V}}f(x,\xi) = \sum_{i=1}^n \beta_i(\xi)\phi_i(x). \quad (8)$$

Here, $\beta_i(\xi) \in \mathbb{R}, i = 1, \dots, n$, are the coefficients corresponding to the basis function $\phi_i(x)$. Since the basis functions $\phi_i(x), i = 1, \dots, n$ are orthonormal, the coefficient β_i can be calculated by

$$\beta_i(\xi) = \langle f(x,\xi), \phi_i(x) \rangle = \int_{\Omega} f(x,\xi)\phi_i(x) dx, \quad \forall i, 1 \leq i \leq n. \quad (9)$$

However, in order to get the coefficient $\beta_i(\xi)$ in (9), we need to evaluate an integration with target function $f(x,\xi)$, which is inconvenient.

The interpolation method [2,18,21] is widely used in function approximation. It approximates the target function $f(x,\xi)$ by restricting the values of this function on n interpolation points $x^{(1)}, \dots, x^{(n)} \in \Omega$. If we denote the approximation function as $\mathcal{I}_{\Omega}f(x,\xi) \approx f(x,\xi)$, the restrictions are

$$\mathcal{I}_{\Omega}f(x^{(i)}, \xi) = f(x^{(i)}, \xi), \quad \forall \xi \in \Gamma, \forall i, 1 \leq i \leq n. \quad (10)$$

In addition, if we approximate the target function $f(x,\xi)$ using interpolation methods in a linear approximation space \mathcal{V} , and denote it as $\mathcal{I}_{\mathcal{V}}f(x,\xi)$, then (8) and (10) can be written as

$$\mathcal{I}_{\mathcal{V}}f(x^{(i)}, \xi) = \sum_{i=1}^n \beta_i(\xi)\phi_i(x^{(i)}) = f(x^{(i)}, \xi), \quad \forall \xi \in \Gamma, \forall i, 1 \leq i \leq n. \quad (11)$$

Here, $\beta_i(\xi), i = 1, \dots, n$, are the unknown coefficients. Moreover, we denote the matrix $\Phi \in \mathbb{R}^{n \times n}$ as

$$\Phi = \begin{pmatrix} \phi_1(x^{(1)}) & \dots & \phi_n(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \phi_1(x^{(n)}) & \dots & \phi_n(x^{(n)}) \end{pmatrix}.$$

Then, the coefficient $\underline{\beta}(\xi) := (\beta_1(\xi), \dots, \beta_n(\xi))^T$ satisfying (11) is the solution of the linear system

$$\begin{aligned} \Phi \underline{\beta}(\xi) &= \begin{pmatrix} \phi_1(x^{(1)}) & \dots & \phi_n(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \phi_1(x^{(n)}) & \dots & \phi_n(x^{(n)}) \end{pmatrix} \begin{pmatrix} \beta_1(\xi) \\ \vdots \\ \beta_n(\xi) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \beta_i(\xi)\phi_i(x^{(1)}) \\ \vdots \\ \sum_{i=1}^n \beta_i(\xi)\phi_i(x^{(n)}) \end{pmatrix} = \begin{pmatrix} f(x^{(1)}, \xi) \\ \vdots \\ f(x^{(n)}, \xi) \end{pmatrix}. \end{aligned} \quad (12)$$

Compared with (9), $\underline{\beta}$ which is evaluated by the interpolation method (12) avoids the integration and only needs to know the value of target function $f(x, \xi)$ on n interpolation points $x^{(1)}, \dots, x^{(n)}$. The target function $f(x, \xi)$ can then be approximated by

$$f(x, \xi) \approx \mathcal{I}_{\mathcal{V}} f(x, \xi) = \sum_{i=1}^n \beta_i(\xi) \phi_i(x), \quad (13)$$

where the coefficients $\beta_i(\xi), i = 1, \dots, n$ are the solution of (12). Then, (13) naturally satisfies the restriction (11) by the definition of the coefficients $\beta_i(\xi), i = 1, \dots, n$.

In EIM [2,3,17,21], systematical approaches to choose suitable interpolation points are given. Here, approximation basis functions and interpolation points are typically obtained alternately by minimizing the error between the target function $f(x, \xi)$ and its approximation function $\mathcal{I}_{\mathcal{V}} f(x, \xi)$, of which the produce can be summarized as follows. First, we initialize $\mathcal{V}_0 = \text{span}(0)$, and select a new parameter value as

$$\xi^{(k+1)} := \operatorname{argmax}_{\xi \in \Gamma_{\text{train}}} \|f(x, \xi) - \mathcal{I}_{\mathcal{V}_k} f(x, \xi)\|.$$

For the parameter value $\xi^{(k+1)}$, if the error

$$\eta(\xi^{(k+1)}) = \|f(x, \xi^{(k+1)}) - \mathcal{I}_{\mathcal{V}_k} f(x, \xi^{(k+1)})\| \quad (14)$$

is greater than tol , the linear approximation space is updated as $\mathcal{V}_{k+1} = \text{span}(\mathcal{V}_k \cup f(x, \xi^{(k+1)}))$. Then, we orthonormalize the target function $f(x, \xi^{(k+1)})$ with the basis functions $\phi_1(x), \dots, \phi_k(x)$ by Schmidt orthogonalization for numerical stability and denote it as $\phi_{k+1}(x)$, i.e.,

$$e(x) = f(x, \xi^{(k+1)}) - \sum_{i=1}^k \langle f(x, \xi^{(k+1)}), \phi_i(x) \rangle \phi_i(x),$$

$$\phi_{k+1}(x) = \frac{e(x)}{\|e(x)\|},$$

where $\langle f(x, \xi^{(k+1)}), \phi_i(x) \rangle, i = 1, 2, \dots, k$, are the coefficients of the projection of $f(x, \xi^{(k+1)})$ over $\mathcal{V}_k = \text{span}(\{\phi_1(x), \phi_2(x), \dots, \phi_k(x)\})$. The next interpolation point is selected by maximizing the absolute value of the error $r_k(x) = \phi_{k+1}(x) - \mathcal{I}_{\mathcal{V}_k} \phi_{k+1}(x)$, i.e.,

$$x^{(k+1)} = \operatorname{argmax}_{x \in \Omega} |r_k(x)|.$$

The above procedure is repeated until the error $\eta(\xi) \leq tol$ (see (14)) for all $\xi \in \Gamma_{\text{train}}$, and we denote $\mathcal{V} = \mathcal{V}_n$ for the final step. The relationship between the error $\|f(x, \xi) - \mathcal{I}_{\mathcal{V}} f(x, \xi)\|$ and $\|f(x, \xi) - \mathcal{P}_{\mathcal{V}} f(x, \xi)\|$ are discussed in [2] (see [6] for its discrete version).

3. Discrete Empirical Interpolation Method and Its Stochastic Version

We evaluate the values of $f(x, \xi^{(k)})$ on a discrete physical domain, that is, computing a vector function

$$\xi \mapsto \mathbf{f}_h(\xi),$$

where $\mathbf{f}_h(\xi) = (f(x^{(1)}, \xi), \dots, f(x^{(N_h)}, \xi))^T$ with N_h components is the discrete version of $f(x, \xi)$ on N_h physical points $x^{(1)}, \dots, x^{(N_h)}$. The number of discrete points N_h is usually large to meet certain appropriation accuracy. On the other hand, since the physical domain has been discretized, selecting interpolation points is to find proper indices from $1, 2, \dots, N_h$. If i_1, \dots, i_n denote the indices of the interpolation points and $\mathbf{P} \in \mathbb{R}^{N_h \times n}$ denote the matrix of interpolation points defined as

$$\mathbf{P} = [e_{i_1}, \dots, e_{i_n}]$$

with $e_i \in \mathbb{R}^{N_h \times 1}$ the i -th canonical unit vector, the values of target function $\mathbf{f}_h(\xi)$ on the interpolation points can be written as $\mathbf{P}^T \mathbf{f}_h(\xi)$. As a result, the corresponding interpolation only needs the values of $\mathbf{f}_h(\xi)$ on i_1, \dots, i_n components with $n \ll N_h$.

3.1. Discrete Empirical Interpolation Method (DEIM)

Before presenting our stochastic discrete empirical interpolation method, we here review the original DEIM [6]. In the discrete formulation, a linear approximation space \mathcal{V} can be expressed as

$$\mathcal{V} = \text{span}(\{\mathbf{f}_h(\xi^{(1)}), \dots, \mathbf{f}_h(\xi^{(n)})\}),$$

which is an approximation of the discrete target manifold

$$\mathcal{M}_h = \{\mathbf{f}_h(\xi) | \xi \in \Gamma\}.$$

Denoting a set of orthonormal basis functions of this linear approximation space \mathcal{V} by $\{\mathbf{q}_i \in \mathbb{R}^{N_h \times 1}, i = 1, 2, \dots, n\}$, and denoting the matrix of basis functions by $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_n] \in \mathbb{R}^{N_h \times n}$, the approximation of the target function $\mathbf{f}_h(\xi)$ corresponding to (8) can be written as

$$\mathbf{f}_h(\xi) \approx \hat{\mathbf{f}}_h := \mathbf{Q}\beta(\xi) = \sum_{i=1}^n \beta_i(\xi) \mathbf{q}_i,$$

where $\beta(\xi) = (\beta_1, \dots, \beta_n)^T$ is the coefficient. The restriction of interpolation points corresponding to (13) is

$$\mathbf{P}^T \mathbf{f}_h(\xi) = \mathbf{P}^T \mathbf{Q} \beta(\xi),$$

which gives $\beta(\xi) = (\mathbf{P}^T \mathbf{Q})^{-1} \mathbf{P}^T \mathbf{f}_h(\xi)$. Then, it can naturally derive

$$\hat{\mathbf{f}}_h(\xi) = \mathbf{Q} \beta(\xi) = \underbrace{\mathbf{Q}(\mathbf{P}^T \mathbf{Q})^{-1}}_{\tilde{\mathbf{Q}}} \underbrace{\mathbf{P}^T \mathbf{f}_h(\xi)}_{\tilde{\beta}(\xi)}. \quad (15)$$

By the way β is obtained, the values of the approximation function $\hat{\mathbf{f}}_h(\xi)$ are equal to the target function $\mathbf{f}_h(\xi)$ on the components i_1, \dots, i_n for any $\xi \in \Gamma$, i.e., $\mathbf{P}^T \hat{\mathbf{f}}_h(\xi) = \mathbf{P}^T \mathbf{f}_h(\xi)$. Moreover, the calculation of the new matrix of basis functions $\tilde{\mathbf{Q}} = \mathbf{Q}(\mathbf{P}^T \mathbf{Q})^{-1} \in \mathbb{R}^{N_h \times n}$ is required once only. For a new realization ξ , DEIM only computes the values of the target function $\mathbf{f}_h(\xi)$ on the components i_1, \dots, i_n as the new coefficient $\tilde{\beta}(\xi) = \mathbf{P}^T \mathbf{f}_h(\xi) \in \mathbb{R}^{n \times 1}$.

The procedure of DEIM to construct a linear approximation space and select interpolation points can be stated as follows. For a fine candidate sample set Γ_{train} , it first chooses

$$\xi^{(1)} = \operatorname{argmax}_{\xi \in \Gamma_{\text{train}}} \|\mathbf{f}_h(\xi)\|_2,$$

where $\|\cdot\|_2$ is the vector L_2 norm for a discrete function. Then, we orthonormalize $\mathbf{f}_h(\xi^{(1)})$ as \mathbf{q}_1 and denote $\mathbf{Q}_1 = [\mathbf{q}_1]$ with

$$\mathbf{q}_1 := \frac{\mathbf{f}_h(\xi^{(1)})}{\|\mathbf{f}_h(\xi^{(1)})\|_2}.$$

The index of the first interpolation point i_1 is initialized as

$$i_1 = \operatorname{argmax}_{i=1, \dots, N_h} |e_i^T \mathbf{q}_1|, \quad (16)$$

and we let $\mathbf{P}_1 = [e_{i_1}]$. For the $(k+1)$ -th basis function and interpolation point, a parameter value $\zeta^{(k+1)}$ is chosen from the candidate sample set Γ_{train} by finding the maximum error between the target function $\mathbf{f}_h(\zeta)$ and its approximation $\hat{\mathbf{f}}_h(\zeta)$ in the linear approximation space $\mathcal{V}_k = \text{span}(\{\mathbf{q}_1, \dots, \mathbf{q}_k\})$, i.e.,

$$\zeta^{(k+1)} = \arg\max_{\zeta \in \Gamma_{\text{train}}} \|\mathbf{f}_h(\zeta) - \hat{\mathbf{f}}_h(\zeta)\|_2. \quad (17)$$

If the error $\eta(\zeta^{(k+1)}) = \|\mathbf{f}_h(\zeta^{(k+1)}) - \hat{\mathbf{f}}_h(\zeta^{(k+1)})\|_2$ is greater than a given tolerance tol , we orthonormalize $\mathbf{f}_h(\zeta^{(k+1)})$ with $\mathbf{q}_1, \dots, \mathbf{q}_k$ by Schmidt orthogonalization and denote it as \mathbf{q}_{k+1} , that is,

$$\begin{aligned} \mathbf{e}^{(k+1)} &:= \mathbf{f}_h(\zeta^{(k+1)}) - \mathbf{Q}_k \mathbf{Q}_k^T \mathbf{f}_h(\zeta^{(k+1)}), \\ \mathbf{q}^{(k+1)} &:= \frac{\mathbf{e}^{(k+1)}}{\|\mathbf{e}^{(k+1)}\|_2}. \end{aligned} \quad (18)$$

Then, the matrix of basis functions is updated as $\mathbf{Q}_{k+1} = [\mathbf{Q}_k, \mathbf{q}_{k+1}]$. The residual of interpolation projecting \mathbf{q}_{k+1} to the linear approximation space \mathcal{V}_k can be written as

$$\mathbf{r}_k := \mathbf{q}_{k+1} - \hat{\mathbf{q}}_{k+1} = \mathbf{q}_{k+1} - \mathbf{Q}_k (\mathbf{P}_k^T \mathbf{Q}_k)^{-1} \mathbf{P}_k^T \mathbf{q}_{k+1}. \quad (19)$$

An index of $(k+1)$ -th interpolation point i_{k+1} is selected by finding the maximum component of residual \mathbf{r}_k , i.e.,

$$i_{k+1} = \arg\max_{i=1, \dots, N_h} |e_i^T \mathbf{r}_k|, \quad (20)$$

and we update $\mathbf{P}_{k+1} = [\mathbf{P}_k, e_{i_{k+1}}]$. Finally, we denote $\mathcal{V} = \mathcal{V}_n$, $\mathbf{Q} = \mathbf{Q}_n$ and $\mathbf{P} = \mathbf{P}_n$. This linear approximation space \mathcal{V} then satisfies

$$\sup_{\zeta \in \Gamma_{\text{train}}} \|\mathbf{f}_h(\zeta) - \hat{\mathbf{f}}_h(\zeta)\|_2 \leq tol.$$

The whole procedure is stated in Algorithm 1, where η means the error in (14) or its discrete form (17).

Algorithm 1 Discrete empirical interpolation method (DEIM) [6]

Input: A candidate sample set Γ_{train} and a target function $\mathbf{f}_h(\zeta)$.

- 1: Initialize $\zeta^{(1)} = \arg\max_{\zeta \in \Gamma_{\text{train}}} \|\mathbf{f}_h(\zeta)\|_2$ and orthonormalize $\mathbf{f}_h(\zeta^{(1)})$ as $\mathbf{q}_1 = \mathbf{f}_h(\zeta^{(1)}) / \|\mathbf{f}_h(\zeta^{(1)})\|_2$.
- 2: Initialize $i_1 = \arg\max_{i=1, \dots, N_h} |e_i^T \mathbf{q}_1|$.
- 3: Initialize $\mathbf{Q} = [\mathbf{q}_1]$ and $\mathbf{P} = [e_{i_1}]$.
- 4: **while** $\sup_{\zeta \in \Gamma_{\text{train}}} \eta(\zeta) = \|\mathbf{f}_h(\zeta) - \hat{\mathbf{f}}_h(\zeta)\|_2 > tol$ **do**
- 5: Compute the error $\eta(\zeta^{(i)})$ for $\zeta^{(i)} \in \Gamma_{\text{train}}, i = 1, 2, \dots, |\Gamma_{\text{train}}|$.
- 6: Let $\zeta^{(k+1)} = \arg\max_{\zeta \in \Gamma_{\text{train}}} \eta(\zeta)$.
- 7: Compute \mathbf{q}_{k+1} through orthonormalizing $\mathbf{f}_h(\zeta^{(k+1)})$ with $\mathbf{q}_1, \dots, \mathbf{q}_k$ by (18).
- 8: Solve the equation $\mathbf{P}^T \mathbf{q}_{k+1} = \mathbf{P}^T \mathbf{Q} \beta$ for β .
- 9: Compute the residual $\mathbf{r}_k = \mathbf{q}_{k+1} - \hat{\mathbf{q}}_{k+1} = \mathbf{q}_{k+1} - \mathbf{Q} \beta$.
- 10: Select the interpolation index i_{k+1} as $i_{k+1} = \arg\max_{i=1, \dots, N_h} |e_i^T \mathbf{r}_k|$.
- 11: Update $\mathbf{Q} = [\mathbf{Q}, \mathbf{q}_{k+1}]$ and $\mathbf{P} = [\mathbf{P}, e_{i_{k+1}}]$.
- 12: **end while**

Output: The matrix of basis functions \mathbf{Q} and the matrix of interpolation points \mathbf{P} .

3.2. Stochastic Discrete Empirical Interpolation Method (SDEIM)

For standard EIM (or DEIM), to satisfy the tol -approximation (see (7)), the candidate sample set Γ_{train} typically needs to be fine enough (i.e., the condition (5) holds). However, if the size of Γ_{train} is large, it can be expensive to construct the DEIM approximation (15). In this section, we propose a stochastic discrete empirical interpolation method (SDEIM). In SDEIM, we accept the approximation with probability instead of giving a threshold with certainty, and we then avoid fine candidate sample sets. We note that a weight EIM strategy is developed in [4], while the purpose of our work is to give a stochastic criterion for updating training sets.

Our problem formulation is still to approximate the target manifold \mathcal{M}_h with a linear approximation space \mathcal{V} , such that it satisfies

$$\eta(\xi) = \|\mathbf{f}_h(\xi) - \mathcal{I}_{\mathcal{V}}\mathbf{f}_h(\xi)\|_2 \leq tol.$$

However, assuming that a probability measure \mathbb{P} exists in the parameter space Γ , we herein do not ensure $\eta(\xi) \leq tol$ for all $\xi \in \Gamma$, but concern about the failure probability

$$p := \mathbb{P}(\xi \in \Gamma \mid \eta(\xi) > tol), \quad (21)$$

which measures the size of the parameter set where the approximation is not accurate enough. While the probability p can hardly be exactly evaluated, we evaluate its empirical probability among N samples

$$\bar{p} := \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\eta(\xi^{(i)}) - tol), \quad (22)$$

where $\mathbb{I}(x)$ is the indicator function defined as

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

The probability (22) is to calculate the average number of occurrences of $\eta(\xi) > tol$ on N samples. By the law of large numbers [22], the empirical probability \bar{p} converges to the probability p with probability one as N goes to infinite, i.e.,

$$\mathbb{P}\left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\eta(\xi^{(i)}) - tol) = p\right) = 1.$$

Since the implicit constant p reflects the probability that ξ does not satisfy the tolerance tol , we hope that p is small enough. On the other hand, p can not be evaluated explicitly, and the empirical probability \bar{p} is an approximation of p . In SDEIM, therefore, \bar{p} is set to be small enough. For convenience, we take $\bar{p} = 0$ in the verifying stage and use the sample size N to control the accuracy of \bar{p} to approximate p .

The procedure of SDEIM has two steps: constructing a linear approximation space \mathcal{V} , and verifying whether the empirical probability $\bar{p} = 0$ for this linear approximation space \mathcal{V} in N consecutive samples. Our SDEIM algorithm can be described as follows. First, a sample $\xi^{(1)} \in \Gamma$ is randomly selected, and the target function $\mathbf{f}_h(\xi^{(1)})$ is normalized as \mathbf{q}_1 . The linear approximation space is initialized as $\mathcal{V}_1 = \text{span}(\{\mathbf{q}_1\})$ and the matrix of basis function is denoted as $\mathbf{Q}_1 = [\mathbf{q}_1]$. The index of the first interpolation point i_1 is initialized in the same way as (16) and the matrix of interpolation point is denoted as $\mathbf{P}_1 = [e_{i_1}]$. Then, the empirical probability \bar{p} is verified for this linear approximation space \mathcal{V}_1 . We sample N consecutive samples. If one of these N samples makes $\eta(\xi) > tol$, we orthonormalize \mathbf{q}_{k+1} for this sample ξ , find the index of the $(k+1)$ -th interpolation point i_{k+1} as DEIM in (19) and (20) and update $\mathbf{P}_{k+1} = [\mathbf{P}_k, e_{i_{k+1}}]$. The linear approximation space is updated as $\mathcal{V}_{k+1} = \text{span}(\mathcal{V}_k \cup \mathbf{q}_{k+1})$ and the matrix of the basis function is updated

as $\mathbf{Q}_{k+1} = [\mathbf{Q}_k, \mathbf{q}_{k+1}]$. Then, the empirical probability \bar{p} is verified again for this linear approximation space \mathcal{V}_{k+1} . The above steps are repeated alternately until there is a linear approximation space \mathcal{V}_n such that $\eta(\xi^{(i)}) \leq tol$ for N consecutive samples. Finally, a linear approximation space $\mathcal{V} = \mathcal{V}_n$ is obtained, such that the empirical probability $\bar{p} = 0$ for N consecutive samples. Details of SDEIM are stated as Algorithm 2.

Algorithm 2 Stochastic discrete empirical interpolation method (SDEIM)

Input: A constant number N and a target function $\mathbf{f}_h(\xi)$.

- 1: Sample ξ randomly.
- 2: Evaluate the target function $\mathbf{f}_h(\xi)$ and initialize $\mathbf{q}_1 = \mathbf{f}_h(\xi) / \|\mathbf{f}_h(\xi)\|_2$.
- 3: Initialize $i_1 = \operatorname{argmax}_{i=1, \dots, N_h} |e_i^T \mathbf{q}_1|$.
- 4: Initialize $\mathbf{Q} = [\mathbf{q}_1]$ and $\mathbf{P} = [\mathbf{e}_{i_1}]$.
- 5: Initialize the counting index $j = 0$.
- 6: **while** $j \leq N$ (if $j > N$, it means that $\bar{p} = 0$ in verifying stage) **do**
- 7: Sample $\xi \in \Gamma$ randomly.
- 8: **if** error $\eta(\xi) = \|\mathbf{f}_h(\xi) - \hat{\mathbf{f}}_h(\xi)\|_2 > tol$ **then**
- 9: Set counting index $j = 0$.
- 10: Compute \mathbf{q}_{k+1} through orthonormalizing $\mathbf{f}_h(\xi)$ with the $\mathbf{q}_1, \dots, \mathbf{q}_k$ by (18).
- 11: Solve $\mathbf{P}^T \mathbf{q}_{k+1} = \mathbf{P}^T \mathbf{Q} \beta$ for β .
- 12: Compute the residual $\mathbf{r}_k = \mathbf{q}_{k+1} - \hat{\mathbf{q}}_{k+1} = \mathbf{q}_{k+1} - \mathbf{Q} \beta$.
- 13: Select the interpolation index i_{k+1} as $i_{k+1} = \operatorname{argmax}_{i=1, \dots, N_h} |e_i^T \mathbf{r}_k|$.
- 14: Update $\mathbf{Q} = [\mathbf{Q}, \mathbf{q}_{k+1}]$ and $\mathbf{P} = [\mathbf{P}, \mathbf{e}_{i_{k+1}}]$.
- 15: **else**
- 16: Update $j = j + 1$.
- 17: **end if**
- 18: **end while**

Output: The matrix of basis functions \mathbf{Q} and the matrix of interpolation points \mathbf{P} .

3.3. Performance and Complexity of SDEIM

In SDEIM, we ensure that the empirical probability $\bar{p} = 0$ for N consecutive samples in the verifying stage, i.e., the number of ξ such that $\eta(\xi) > tol$ is zero in these N consecutive samples. The probability p is adjusted through the samples size N , which affects the accuracy of \bar{p} for approximating p . For any tolerance tol and sample size N , there is always a linear approximation space \mathcal{V} such that $\eta(\xi) = \|\mathbf{f}_h(\xi) - \mathcal{I}_{\mathcal{V}} \mathbf{f}_h(\xi)\|_2 \leq tol$ for N consecutive samples—in the worst case, when the dimension of the linear approximation space $\dim(\mathcal{V}) = N_h$, this linear approximation space \mathcal{V} obviously satisfies the condition.

Although the probability value p cannot be evaluated explicitly, the empirical probability \bar{p} approximates the probability p with probability one as N goes to infinite. Hence, for a given threshold ε and a confidence $(1 - \delta)$, we can consider this question, whether there is

$$\mathbb{P}(|p - \bar{p}| < \varepsilon) \geq 1 - \delta. \quad (23)$$

By the law of large numbers, the answer is always correct for a suitable N . Before describing the relationship between N and the probability p , the threshold ε and the confidence level $(1 - \delta)$, we first introduce the Hoeffding's inequality as Lemma 1.

Lemma 1 (Hoeffding's inequality [23]). *Let random variables X_1, X_2, \dots, X_N be independent and identically distributed with values in the interval $[0, 1]$ and expectation EX , then for any $\varepsilon > 0$*

$$\begin{aligned} \mathbb{P}\left(\left|EX - \frac{1}{N} \sum_{i=1}^N X_i\right| \geq \varepsilon\right) &\leq 2 \exp(-2N\varepsilon^2), \\ \mathbb{P}\left(EX - \frac{1}{N} \sum_{i=1}^N X_i \geq \varepsilon\right) &\leq \exp(-2N\varepsilon^2). \end{aligned}$$

The Hoeffding's inequality characterizes the relationship between the arithmetic mean $\frac{1}{N} \sum_{i=1}^N X_i$ and its expectation $\mathbb{E}X$ for bounded random variables. In SDEIM, when a linear approximation space \mathcal{V} is given, the error $\eta(\xi)$ can only be greater than tolerance tol or not greater than tol for any realization ξ . Hence, the indicator $\mathbb{I}(\eta(\xi) - tol)$ is a random variable taking value of zero or one with expectation $\mathbb{E}\mathbb{I}(\eta(\xi) - tol) = p$. Note that we set the empirical probability $\bar{p} = 0$ in the verifying stage. Then, the problem (23) can be answered with Theorem 1.

Theorem 1. For any significant level δ ($0 < \delta < 1$) and threshold ε , the linear approximation space \mathcal{V} is produced by SDEIM with sample size N . Then, if $N \geq \frac{1}{2\varepsilon^2} \ln \frac{1}{\delta}$, we have

$$\mathbb{P}(p < \varepsilon) \geq 1 - \delta.$$

Proof. By Lemma 1 and $\bar{p} = 0$ in SDEIM, we have

$$\begin{aligned} \mathbb{P}(p < \varepsilon) &= 1 - \mathbb{P}(p \geq \varepsilon) \\ &= 1 - \mathbb{P}(p - \bar{p} \geq \varepsilon) \\ &\geq 1 - \exp(-2N\varepsilon^2) \\ &\geq 1 - \exp(-2\varepsilon^2 \frac{1}{2\varepsilon^2} \ln \frac{1}{\delta}) \\ &= 1 - \delta. \end{aligned}$$

□

Moreover, the relationship between probability p and the sample size N , the threshold ε , the confidence $(1 - \delta)$, can be explicitly described as Theorem 2.

Theorem 2. For any significance level δ ($0 < \delta < 1$), the linear approximation space \mathcal{V} is produced by SDEIM with sample size N . There is at least confidence $(1 - \delta)$ satisfying

$$0 < p < \sqrt{\frac{1}{2N} \ln \frac{1}{\delta}}.$$

It means that, if the error estimator in SDEIM is set to $\eta(\xi) = \|\mathbf{f}_h(\xi) - \mathcal{I}_{\mathcal{V}}\mathbf{f}_h(\xi)\|_2$, with confidence $(1 - \delta)$, we have

$$\mathbb{P}\left(\|\mathbf{f}_h(\xi) - \mathcal{I}_{\mathcal{V}}\mathbf{f}_h(\xi)\|_2 > tol\right) < \sqrt{\frac{1}{2N} \ln \frac{1}{\delta}} \quad (24)$$

and

$$\mathbb{P}\left(\|\mathbf{f}_h(\xi) - \mathcal{I}_{\mathcal{V}}\mathbf{f}_h(\xi)\|_2 \leq tol\right) \geq 1 - \sqrt{\frac{1}{2N} \ln \frac{1}{\delta}}. \quad (25)$$

Proof. Without loss of generality, we set

$$\delta = \exp(-2N\varepsilon^2)$$

for suitable ε , that is

$$\varepsilon = \sqrt{\frac{1}{2N} \ln \frac{1}{\delta}}.$$

By Lemma 1 and $\bar{p} = 0$ in SDEIM, we have

$$\mathbb{P}\left(p \geq \sqrt{\frac{1}{2N} \ln \frac{1}{\delta}}\right) \leq \exp(-2N\varepsilon^2) = \delta.$$

That is

$$p < \sqrt{\frac{1}{2N} \ln \frac{1}{\delta}}$$

with confidence $(1 - \delta)$. In addition, when the error estimator is set to $\eta(\xi) = \|\mathbf{f}_h(\xi) - \mathcal{I}_V \mathbf{f}_h(\xi)\|_2$, since the definition of p in (21) is

$$p = \mathbb{P}(\xi \in \Gamma \mid \eta(\xi) > tol),$$

(24) and (25) can be derived directly. \square

Theorem 2 ensures that the linear approximation space \mathcal{V} given by the SDEIM algorithm (Algorithm 2), can approximate the target function $\mathbf{f}_h(\xi)$ with the tolerance tol and the probability threshold ε . Actually, in Algorithm 2, we reset the counting index each time $\eta(\xi) > tol$ occurs, which means that there are more than N samples such that $\eta(\xi) < tol$. Hence, $\sqrt{\frac{1}{2N} \ln \frac{1}{\delta}}$ is a more strict upper bound.

In the procedure of Algorithm 2, the best situation is that n consecutive samples are compared to generate the linear approximation space \mathcal{V} with dimension $\dim(\mathcal{V}) = n$ and then N samples are compared to verify the error $\eta(\xi) \leq tol$. In this case, the number of comparisons is $\mathcal{O}(n + N)$. The worst situation is that, each sample ξ is found with $\eta(\xi) > tol$ at the end of the N comparisons for verification and this procedure is repeated n times. It is clear that, the number of comparisons in SDEIM is much smaller than that in standard DEIM, where large training sets typically need to be looped over.

4. Numerical Experiments

In this section, four test problems are considered to show the efficiency of SDEIM. The first one is a nonlinear parameterized function with spatial points in one dimension. The second one is to extend the first experiment to two dimensions. The third one focuses on the property of SDEIM for random fields. The last experiment is a nonlinear ordinary differential equation arising in neuron modeling.

4.1. A Nonlinear Parameterized Function with Spatial Points in One Dimension

Consider a nonlinear parameterized function $f : \Omega \times \Gamma \rightarrow \mathbb{R}$ defined by

$$f(x, \xi) = 10x \sin(2\pi\xi x),$$

where $x \in \Omega = [0, 1] \subset \mathbb{R}$ and $\xi \in \Gamma = [1, T] \subset \mathbb{R}$ for a constant T . For a given parameter ξ , the function $f(x, \xi)$ has the period $1/\xi$. Figure 1 plots the function $f(x, \xi)$ for different $\xi = 1, 2$ and 8 . It shows that the target function can have different complexities for different parameter ranges. Let $\{x^{(i)}, i = 1, 2, \dots, N_h\}$ be a uniform grid points in Ω for $N_h = 400$, and define $\mathbf{f}_h(\xi) : \Gamma \rightarrow \mathbb{R}^{N_h}$ as

$$\mathbf{f}_h(\xi) = \left[f(x^{(1)}, \xi), \dots, f(x^{(N_h)}, \xi) \right]^T \in \mathbb{R}^{N_h}$$

for $\xi \in \Gamma$. Let the range of the parameter be $\Gamma = [1, T]$ for $T = 2$ and let Γ_{train} be selected uniformly over Γ with $|\Gamma_{\text{train}}| = 50$ for DEIM. The tolerance tol is set to $tol = 10^{-4}$ and the confidence is set to 0.99 ($\delta = 10^{-2}$) for Algorithms 1 and 2.

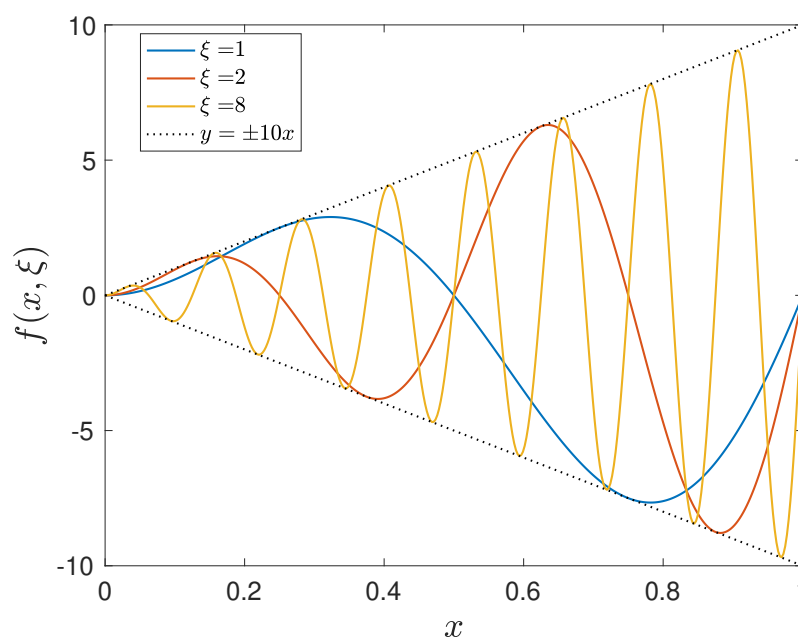


Figure 1. The parameterized function $f(x, \xi)$ for $\xi = 1, 2$ and 8 .

Figure 2 shows the training procedures for different methods. Figure 2a is the training procedure of SDEIM with threshold $\varepsilon = 0.70$. An epoch in Figure 2a is a training procedure for a new realization ξ . The black curve indicates that the error changes in the training procedure. The green point is the first successful sample with $\eta(\xi) \leq \text{tol}$. In DEIM, after traversing the candidate sample set, if the sample with the largest error is less than tolerance tol , the algorithm stops and the linear approximation space satisfies $\eta(\xi) \leq \text{tol}$ for all $\xi \in \Gamma_{\text{train}}$. In SDEIM, we continue to find whether there are failed samples (the red crosses) in Γ with $\eta(\xi) > \text{tol}$. The SDEIM algorithm (Algorithm 2) stops when the samples satisfy $\eta(\xi) \leq \text{tol}$ for N consecutive samples. In this case, the probability for the appearance of a failed sample is considered to be very low in SDEIM. Figure 2b shows the relationship between the number of comparisons and the average error in the training procedure for SDEIM (with $p < 0.70$ and $p < 0.10$) and DEIM, where the average error is computed on $\tilde{N} = 10^4$ samples defined as

$$\bar{\eta} := \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \eta(\xi^{(i)}). \quad (26)$$

Since a basis function is found after searching the candidate sample set Γ_{train} in DEIM, its average error decreases in a ladder shape. In SDEIM, errors for an extra N consecutive samples need to be computed, and its numbers of comparisons have a long flat tail.

Table 1 shows more details for the linear approximation space \mathcal{V} produced by the two methods. The average errors $\bar{\eta}$ for different methods are computed using (26) with $\tilde{N} = 10^4$ samples and the empirical probabilities \bar{p} are calculated using (22) with the same $\tilde{N} = 10^4$ samples which are the approximation of p . The number of comparisons is the times in different methods to compare the error $\eta(\xi)$. From Table 1, it can be seen that since SDEIM does not need to search the sample ξ in a candidate sample set Γ_{train} , it has fewer comparisons.

Figure 3a,b show the first six basis functions (i.e., the first six columns of matrix $\tilde{\mathbf{Q}}$) for DEIM and SDEIM. Figure 3c shows the samples ξ which generate the basis functions for DEIM and SDEIM. The black dots in Figure 3c are the samples in the candidate sample set Γ_{train} , and the black dots circled by the blue circles are the samples selected by DEIM. The numbers above them are the order in which they are selected. The red stars are the samples selected by SDEIM, which are generated in random order. Note that the consecutive

samples in DEIM are separated further because they are selected after searching the fine candidate sample set Γ_{train} .

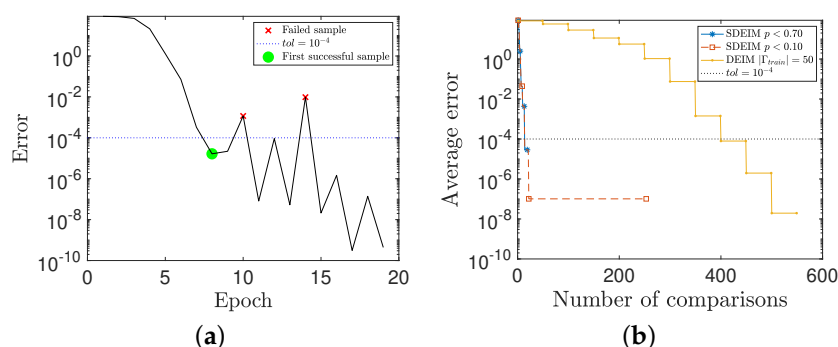


Figure 2. (a) The training procedure of SDEIM with probability threshold $\varepsilon = 0.70$, i.e., $p < 0.70$, where each epoch is to test a new sample ξ for updating the basis. (b) The relationship between the number of comparisons and the average error in the training procedure for SDEIM (with $p < 0.70$ and $p < 0.10$) and DEIM.

Table 1. The average error $\bar{\eta}$, the number of basis functions n for the approximation space \mathcal{V} , the empirical probability \bar{p} and the number of comparisons for SDEIM (with $p < 0.70$ and $p < 0.10$) and DEIM (with $|\Gamma_{\text{train}}| = 50$).

		$\bar{\eta}$	n	\bar{p}	Number of Comparisons
SDEIM	$p < 0.70$	2.8815×10^{-5}	9	7.36×10^{-2}	$14 + N(N = 5)$
	$p < 0.10$	1.0145×10^{-7}	10	0	$22 + N(N = 231)$
DEIM	$ \Gamma_{\text{train}} = 50$	1.9514×10^{-8}	10	0	550

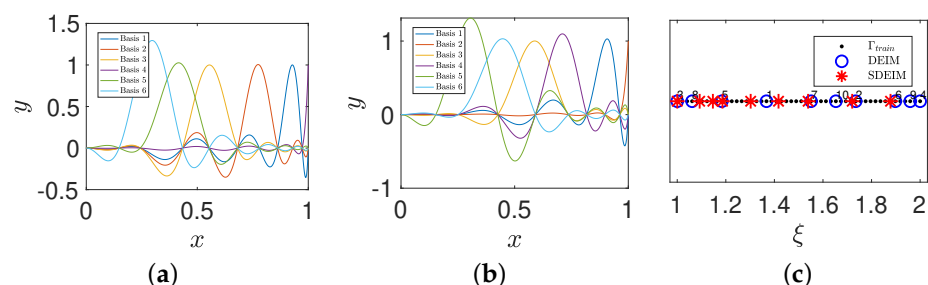


Figure 3. (a) The first six basis functions for DEIM. (b) The first six basis functions for SDEIM. (c) The samples for generating the basis functions.

4.2. A Nonlinear Parameterized Function with Spatial Points in Two Dimensions

Consider a nonlinear parameterized function $f : \Omega \times \Gamma \rightarrow \mathbb{R}$ defined by

$$f(x, \xi) = 10x_1x_2 \sin(2\pi\xi_1x_1) \cos(2\pi\xi_2x_2),$$

where $x = (x_1, x_2) \in \Omega = [0, 1]^2 \subset \mathbb{R}^2$ and $\xi = (\xi_1, \xi_2) \in \Gamma = [1, T]^2 \subset \mathbb{R}^2$ for a constant T . Similarly to Section 4.1, for a large ξ_1 (or ξ_2), $f(x, \xi)$ changes more quickly in the x_1 (or x_2) direction. Let $\{x^{(i)}, i = 1, 2, \dots, N_h\}$ be a uniform grid in Ω for $N_h = 50 \times 50 = 2500$, and define $\mathbf{f}_h(\xi) : \Gamma \rightarrow \mathbb{R}^{N_h}$ as

$$\mathbf{f}_h(\xi) = [f(x^{(1)}, \xi), \dots, f(x^{(N_h)}, \xi)]^T \in \mathbb{R}^{N_h}$$

for $\xi \in \Gamma$. Let the range of the parameter be $\Gamma = [1, T]^2$ for $T = 4$. Figure 4 shows the image of $f(x, \xi)$ for $\xi = [4, 4]$. The tolerance tol is set to $tol = 10^{-4}$ and the confidence is set to 0.99 for $\delta = 10^{-2}$ for Algorithms 1 and 2.

Figure 5 shows the training procedures for the two methods. Where Figure 5a is the training procedure of SDEIM with threshold $\varepsilon = 0.10$. An epoch in Figure 5a is a comparison of the error for a new realization of ξ . The trend of black dots is the trend of errors, which decrease rapidly at early epochs. For these early epochs, the linear approximation space is updated for each realization ξ in this stage. After the first successful sample (the green point with $\eta(\xi) \leq \text{tol}$), the error becomes relatively stable near the tol . Then, for SDEIM, the linear approximation space is updated only for the failed samples (the red crosses with $\eta(\xi) > \text{tol}$), and Algorithm 2 stops when the samples satisfy $\eta(\xi) \leq \text{tol}$ for N consecutive samples. Figure 5b shows the relationship between the number of comparisons and the average error among $\tilde{N} = 10^4$ samples in the training procedure for SDEIM and DEIM. Compared with DEIM, SDEIM has fewer comparisons.

Table 2 shows more details for DEIM and SDEIM with different parameters, where the average errors $\bar{\eta}$ and the empirical probability \bar{p} for different parameters are computed using (22) and (26) on the same $\tilde{N} = 10^4$ samples. It can be seen that the empirical probability \bar{p} for DEIM is similar to that of SDEIM. Moreover, by comparing DEIM with the size of candidate sample set $|\Gamma_{\text{train}}| = 400$ and SDEIM with probability threshold $\varepsilon = 0.10$, the number of comparisons for SDEIM is far less than DEIM.

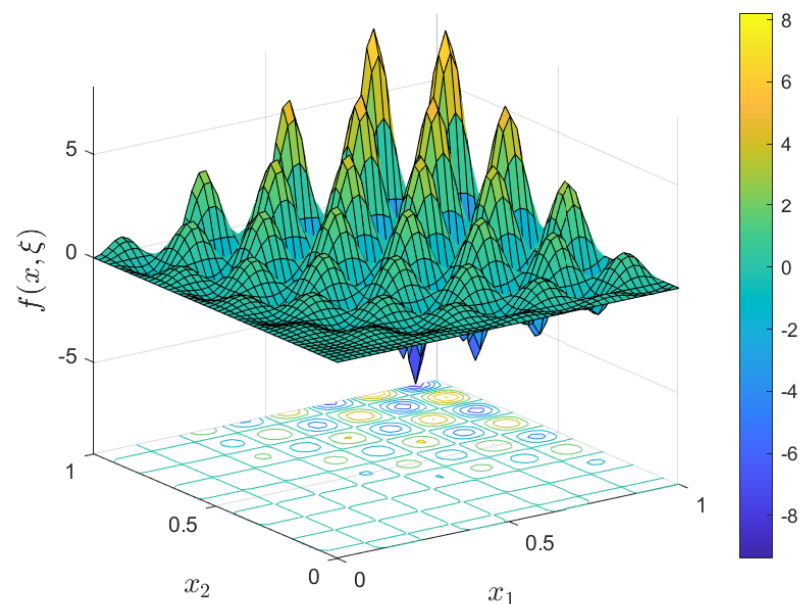


Figure 4. The parameterized function $f(x, \xi)$ for $\xi = [4, 4]$.

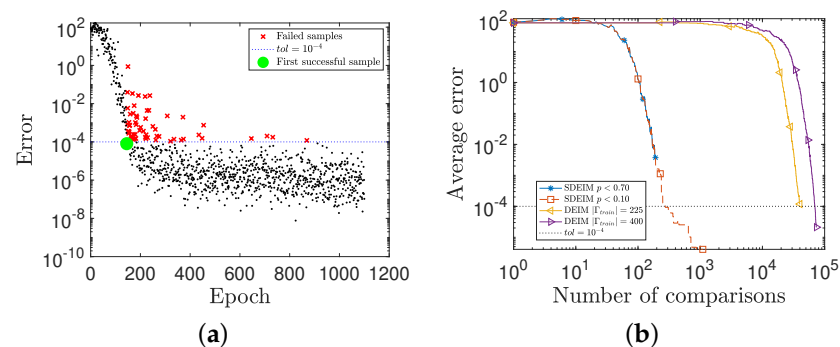


Figure 5. (a) The training procedure of SDEIM with probability threshold $\varepsilon = 0.10$, i.e., $p < 0.10$, where each epoch is to test a new sample ξ for updating the basis. (b) The relationship between the number of comparisons and the average error in the training procedure for DEIM and SDEIM with different parameters.

Table 2. The average error $\bar{\eta}$, the number of basis functions n for the approximation space \mathcal{V} , the empirical probability \bar{p} and the number of comparisons for SDEIM with ($p < 0.70$ and $p < 0.10$) and DEIM with ($|\Gamma_{\text{train}}| = 225$ and $|\Gamma_{\text{train}}| = 400$).

		$\bar{\eta}$	n	\bar{p}	Number of Comparisons
SDEIM	$p < 0.70$	3.7000×10^{-3}	173	2.631×10^{-1}	$186 + N(N = 5)$
	$p < 0.10$	4.0844×10^{-6}	204	2.500×10^{-3}	$869 + N(N = 231)$
DEIM	$ \Gamma_{\text{train}} = 225$	1.1881×10^{-4}	177	1.876×10^{-1}	40,050
	$ \Gamma_{\text{train}} = 400$	2.1180×10^{-5}	184	2.520×10^{-2}	74,000

4.3. Random Fields

Consider a nonlinear parameterized function $f : \Omega \times \Gamma \rightarrow \mathbb{R}$ defined by

$$f(x, \xi) = \exp(z(x, \xi)),$$

where $x = (x_1, x_2) \in \Omega = [0, 1]^2 \subset \mathbb{R}^2$ and $z(x, \xi)$ is assumed to be a stochastic process with mean function $z_0(x) \equiv 1$ and covariance function $\text{Cov}(x, y)$ defined as

$$\text{Cov}(x, y) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|}{L} - \frac{|x_2 - y_2|}{L}\right). \quad (27)$$

Here $x = (x_1, x_2) \in \Omega, y = (y_1, y_2) \in \Omega$ and L is the correlation length. The Karhunen–Loève (KL) expansion (see [24,25] for details) gives a representation of $z(x, \xi)$ as

$$z(x, \xi) = z_0(x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} \varphi_i(x) \xi_i,$$

where $\{\varphi_i(x), i = 1, 2, \dots\}$ are the orthonormal eigenfunctions and $\{\lambda_i, i = 1, 2, \dots\}$ are the corresponding eigenvalues of the covariance function $\text{Cov}(x, y)$. Moreover, $\{\xi_i, i = 1, 2, \dots\}$ are mutually uncorrelated random variables. In this example, we truncate the expansion to M terms as our surrogate model according to the retaining of the energy for $z(x, \xi)$:

$$z_M(x, \xi) := z_0(x) + \sum_{i=1}^M \sqrt{\lambda_i} \varphi_i(x) \xi_i,$$

where $\xi = (\xi_1, \dots, \xi_M)$ is the parameter in the surrogate model whose distribution is assumed to be a uniform distribution in $\Gamma = [-1, 1]^M \subset \mathbb{R}^M$ and M satisfies

$$\delta_{KL} := \frac{\sum_{i=1}^M \lambda_i}{|\Omega| \sigma^2} > 0.95.$$

Here, $|\Omega|$ is the area of the physical domain Ω and the standard deviation σ is set to $\sigma = 0.5$ for (27). We focus on the cases $L = 1$ and $L = 0.5$ in this experiment, and the corresponding dimensions of the parameter spaces are $M = 33$ and $M = 109$ respectively. The physical grid is set to a uniform grid with $N_h = 50 \times 50 = 2500$. We consider the nonlinear parameterized function $\mathbf{f}_h(\xi) : \Gamma \rightarrow \mathbb{R}^{N_h}$ defined by

$$\mathbf{f}_h(\xi) = \left[f_M(x^{(1)}, \xi), \dots, f_M(x^{(N_h)}, \xi) \right]^T \in \mathbb{R}^{N_h}$$

for $\xi \in \Gamma$, where $f_M(x, \xi) = \exp(z_M(x, \xi))$.

Figure 6a,b show the error for SDEIM with dimensions $M = 33$ and $M = 109$ for $\tilde{N} = 10^3$ samples respectively. The tolerance is set to $\text{tol} = 10^{-3}$, the threshold is set to $\varepsilon = 0.30$ and the confidence is set to 0.99 for $\delta = 10^{-2}$ in Algorithm 2. Here the black points are the samples satisfying $\eta(\xi) \leq \text{tol}$ and the red crosses are the failed samples satisfying $\eta(\xi) > \text{tol}$. The empirical probabilities \bar{p} are $\bar{p} = 0.062$ and $\bar{p} = 0.051$, which are both smaller than our probability threshold $\varepsilon = 0.30$.

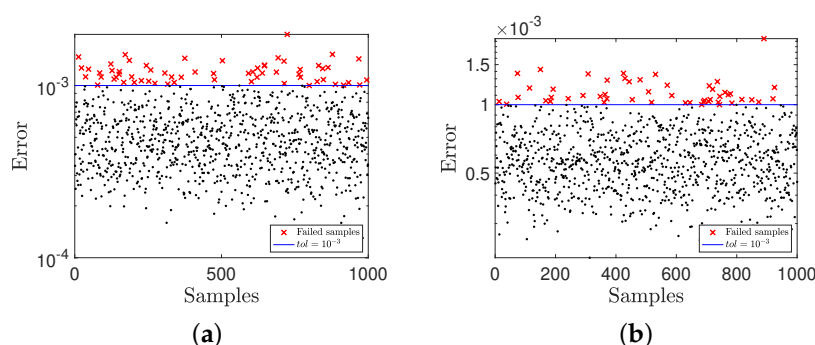


Figure 6. (a) The error for SDEIM with $M = 33$. (b) The error for SDEIM with $M = 109$. Each point or fork corresponds to a sample of the parameter.

Table 3 shows more details about the property of SDEIM in different parameter settings. The data in Table 3 are calculated with $\tilde{N} = 10^4$ samples and the data in Figure 6 are the first 10^3 of these 10^4 samples. From Table 3, it can be seen that for different dimensions of parameter spaces ($M = 33$ and $M = 109$) and different tolerances ($tol = 10^{-1}, 10^{-2}, 10^{-3}$), SDEIM can produce a linear approximation space \mathcal{V} satisfying different probability thresholds ($\varepsilon = 0.70, 0.50, 0.30, 0.10$). Table 4 shows the average number of comparisons for each basis function in SDEIM with different parameter settings. The average number of comparisons is the ratio of the number of comparisons to the number of basis functions for the linear approximation space \mathcal{V} . It is clear that SDEIM has a very small number of comparisons to generate each basis function.

Table 3. The empirical probability \bar{p} for different parameter settings in SDEIM.

		$\varepsilon = 0.70$	$\varepsilon = 0.50$	$\varepsilon = 0.30$	$\varepsilon = 0.10$
		$N = 5$	$N = 10$	$N = 26$	$N = 231$
$M = 33$	$tol = 10^{-1}$	3.060×10^{-1}	2.189×10^{-1}	4.770×10^{-2}	2.700×10^{-3}
	$tol = 10^{-2}$	2.766×10^{-1}	1.550×10^{-1}	6.420×10^{-2}	6.700×10^{-3}
	$tol = 10^{-3}$	5.344×10^{-1}	2.406×10^{-1}	5.550×10^{-2}	1.890×10^{-2}
$M = 109$	$tol = 10^{-1}$	5.664×10^{-1}	2.567×10^{-1}	1.173×10^{-1}	1.050×10^{-2}
	$tol = 10^{-2}$	4.233×10^{-1}	2.143×10^{-1}	4.740×10^{-2}	9.900×10^{-3}
	$tol = 10^{-3}$	1.863×10^{-1}	1.403×10^{-1}	5.020×10^{-2}	4.400×10^{-3}

Table 4. The average number of comparisons for each basis function in SDEIM with different parameter settings.

		$\varepsilon = 0.70$	$\varepsilon = 0.50$	$\varepsilon = 0.30$	$\varepsilon = 0.10$
$M = 33$	$tol = 10^{-1}$	1.1037	1.1841	1.7954	5.4669
	$tol = 10^{-2}$	1.0612	1.1284	1.3482	2.8485
	$tol = 10^{-3}$	1.0346	1.1065	1.5488	2.5069
$M = 109$	$tol = 10^{-1}$	1.0372	1.1270	1.2069	2.6813
	$tol = 10^{-2}$	1.0172	1.0527	1.1066	1.8194
	$tol = 10^{-3}$	1.0291	1.0446	1.1196	2.2821

4.4. The FitzHugh–Nagumo (F–N) System

This test problem considers the F–N system, which is a simplified model of spiking neuron activation and deactivation dynamics [6,16]. Within the F–N system, the nonlinear function $f(v)$ is defined as

$$f(v) = v(v - 0.01)(1 - v), \quad (28)$$

where v is the voltage and satisfies the F-N system

$$\begin{aligned}\varepsilon v_t(x, t) &= \varepsilon^2 v_{xx}(x, t) + f(v(x, t)) - w(x, t) + c, \\ w_t(x, t) &= bv(x, t) - \gamma w(x, t) + c,\end{aligned}\quad (29)$$

with $\varepsilon = 0.015$, $b = 0.5$, $\gamma = 2$, $c = 0.05$ and $x \in \Omega = [0, L]$, $t \geq 0$. The variable w represents the recovery of voltage, and the variable L is set to $L = 1$. Following the settings in [6], initial and boundary conditions are set to

$$\begin{aligned}v(x, 0) &= 0, & w(x, 0) &= 0, & x \in \Omega &= [0, L], \\ v_x(0, t) &= -i_0(t), & v_x(L, t) &= 0, & t \geq 0,\end{aligned}\quad (30)$$

where the stimulus $i_0(t) = 50,000t^3 \exp(-15t)$. We discretize the physical domain Ω using a uniform grid with $N_h = 2500$. The dimension of the finite difference system is 2500. We take 2501 time nodes at evenly in the interval $[0, 8]$, of which 500 are randomly selected as the candidate sample set Γ_{train} for DEIM and SDEIM training, and the remainder are utilized to test the probability features.

The solution of the F-N system has a limit cycle for each spatial variable x , and we display the phase-space diagram of v and w at various spatial positions in Figure 7. Symmetry in the limit cycles of the F-N system is well captured by SDEIM, as seen in Figure 7. Tolerances are set to $\text{tol} = 10^{-3}$ for solutions v, w and $\text{tol} = 10^{-4}$ for nonlinear function $f(v)$. The probability of errors exceeding the corresponding tolerances among 1500 verifying samples are depicted in Figure 8a. The probabilities of v converge to a constant value in this case, because the accuracy of the solution v is dependent on the accuracy of the solution w for both DEIM and SDEIM. Figure 8b shows the number of comparisons increases as the number of basis functions increases. The black dotted line represents the number of comparisons in DEIM, which is $(2|\Gamma_{\text{train}}| - n + 1)n/2$ (n is the number of the basis functions). It is clear that SDEIM is very efficient solve this F-N system.

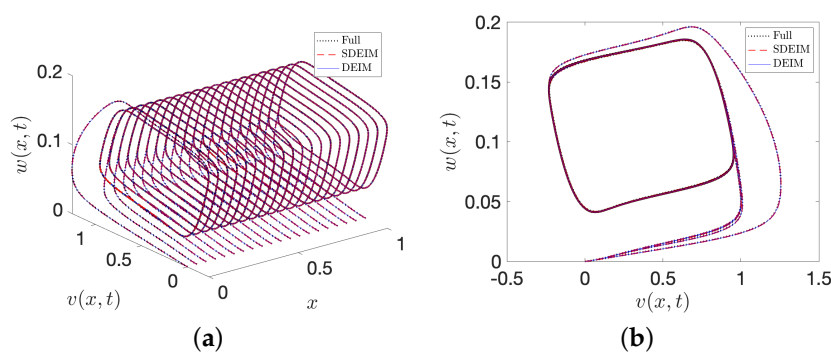


Figure 7. (a) Phase-space diagram of v and w at different spatial points x from the FD system ($\text{dim}2500$), DEIM reduced system ($\text{dim}67$) and SDEIM reduced system ($\text{dim}72$). (b) Corresponding projection of the solution onto the $v - w$ plane.

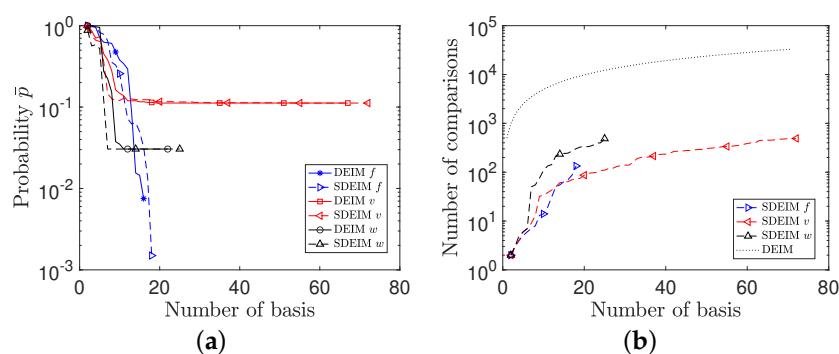


Figure 8. (a) Probability properties for DEIM reduced system and SDEIM reduced system. (b) The number of comparisons for DEIM reduced system and SDEIM reduced system.

5. Conclusions

Empirical interpolation is a widely used model reduction method for nonlinear and nonaffine parameterized systems through a separate representation of the spatial and the parametric variables. This kind of method typically requires a fine candidate sample set, which can lead to high computational costs. With a focus on randomized computational methods, we in this paper propose a stochastic discrete empirical interpolation method (SDEIM). In SDEIM, large candidate sample sets are replaced by gradually generated random samples, such that the computational costs of constructing the corresponding interpolation formulation can be dramatically reduced. With our analysis, the stopping criterion based on a probability threshold in SDEIM can guarantee the interpolation is accurate with a given confidence. Our numerical results show that this randomized approach is efficient, especially for variable separation for high dimensional nonlinear and nonaffine systems. However, as we use the Hoeffding's inequality to estimate the failure probability in the verifying stage, SDEIM is efficient when this probability is not too small, but it can be inefficient when the probability is very small. For applying SDEIM to systems which require high reliability, current efforts are focused on combining subset simulation methods, and implementing such strategies will be the focus of our future work.

Author Contributions: Conceptualization, D.C. and Q.L.; methodology, D.C. and Q.L.; software, D.C. and C.Y.; validation, D.C.; writing—original draft preparation, D.C. and C.Y.; writing—review and editing, D.C. and Q.L.; supervision, Q.L.; project administration, Q.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Science and Technology Commission of Shanghai Municipality (No. 20JC1414300) and the Natural Science Foundation of Shanghai (No. 20ZR1436200).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Benner, P.; Gugercin, S.; Willcox, K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **2015**, *57*, 483–531. [\[CrossRef\]](#)
2. Barrault, M.; Maday, Y.; Nguyen, N.C.; Patera, A.T. An ‘empirical interpolation’ method: Application to efficient reduced-basis discretization of partial differential equations. *C. R. Math.* **2004**, *339*, 667–672. [\[CrossRef\]](#)
3. Maday, Y.; Mula, O. A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation. In *Analysis and Numerics of Partial Differential Equations*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 221–235.
4. Chen, P.; Quarteroni, A.; Rozza, G. A weighted empirical interpolation method: A priori convergence analysis and applications. *ESAIM Math. Model. Numer. Anal.* **2014**, *48*, 943–953. [\[CrossRef\]](#)
5. Elman, H.C.; Forstall, V. Numerical solution of the parameterized steady-state Navier–Stokes equations using empirical interpolation methods. *Comput. Methods Appl. Mech. Eng.* **2017**, *317*, 380–399. [\[CrossRef\]](#)

6. Chaturantabut, S.; Sorensen, D.C. Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.* **2010**, *32*, 2737–2764. [[CrossRef](#)]
7. Peherstorfer, B.; Butnaru, D.; Willcox, K.; Bungartz, H.J. Localized discrete empirical interpolation method. *SIAM J. Sci. Comput.* **2014**, *36*, A168–A192. [[CrossRef](#)]
8. Li, Q.; Jiang, L. A novel variable-separation method based on sparse and low rank representation for stochastic partial differential equations. *SIAM J. Sci. Comput.* **2017**, *39*, A2879–A2910. [[CrossRef](#)]
9. Veroy, K.; Rovas, D.V.; Patera, A.T. A posteriori error estimation for reduced-basis approximation of parametrized elliptic coercive partial differential equations: “Convex inverse” bound conditioners. *ESAIM Control. Optim. Calc. Var.* **2002**, *8*, 1007–1028. [[CrossRef](#)]
10. Quarteroni, A.; Rozza, G. Numerical solution of parametrized Navier–Stokes equations by reduced basis methods. *Numer. Methods Partial Differ. Equ. Int. J.* **2007**, *23*, 923–948. [[CrossRef](#)]
11. Chen, P.; Quarteroni, A.; Rozza, G. Comparison between reduced basis and stochastic collocation methods for elliptic problems. *J. Sci. Comput.* **2014**, *59*, 187–216. [[CrossRef](#)]
12. Jiang, J.; Chen, Y.; Narayan, A. A goal-oriented reduced basis methods-accelerated generalized polynomial chaos algorithm. *SIAM/ASA J. Uncertain. Quantif.* **2016**, *4*, 1398–1420. [[CrossRef](#)]
13. Elman, H.C.; Liao, Q. Reduced basis collocation methods for partial differential equations with random coefficients. *SIAM/ASA J. Uncertain. Quantif.* **2013**, *1*, 192–217. [[CrossRef](#)]
14. Liao, Q.; Lin, G. Reduced basis ANOVA methods for partial differential equations with high-dimensional random inputs. *J. Comput. Phys.* **2016**, *317*, 148–164. [[CrossRef](#)]
15. Cohen, A.; Dahmen, W.; DeVore, R.; Nichols, J. Reduced basis greedy selection using random training sets. *ESAIM Math. Model. Numer. Anal.* **2020**, *54*, 1509–1524. [[CrossRef](#)]
16. Rocsoreanu, C.; Georgescu, A.; Giurgiteanu, N. *The FitzHugh–Nagumo Model: Bifurcation and Dynamics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 10.
17. Grepl, M.A.; Maday, Y.; Nguyen, N.C.; Patera, A.T. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM Math. Model. Numer. Anal.* **2007**, *41*, 575–605. [[CrossRef](#)]
18. Maday, Y.; Stamm, B. Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. *SIAM J. Sci. Comput.* **2013**, *35*, A2417–A2441. [[CrossRef](#)]
19. Temlyakov, V.N. Nonlinear Kolmogorov widths. *Math. Notes* **1998**, *63*, 785–795. [[CrossRef](#)]
20. Cuong, N.N.; Veroy, K.; Patera, A.T. Certified real-time solution of parametrized partial differential equations. In *Handbook of Materials Modeling*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 1529–1564.
21. Kristoffersen, S. The Empirical Interpolation Method. Master’s Thesis, Institutt for Matematiske Fag, Trondheim, Norway, 2013.
22. Dudley, R.M. *Real Analysis and Probability*; CRC Press: Boca Raton, FL, USA, 2018.
23. Hoeffding, W. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*; Springer: Berlin/Heidelberg, Germany, 1994; pp. 409–426.
24. Ghanem, R.G.; Spanos, P.D. *Stochastic Finite Elements: A Spectral Approach*; Courier Corporation: Chelmsford, MA, USA, 2003.
25. Lord, G.J.; Powell, C.E.; Shardlow, T. *An Introduction to Computational Stochastic PDEs*; Cambridge University Press: Cambridge, UK, 2014; Volume 50.