

GRADIENT-PRESERVING HYPER-REDUCTION OF NONLINEAR DYNAMICAL SYSTEMS VIA DISCRETE EMPIRICAL INTERPOLATION*

CECILIA PAGLIANTINI[†] AND FEDERICO VISMARÀ[†]

Abstract. This work proposes a hyper-reduction method for nonlinear parametric dynamical systems characterized by gradient fields such as Hamiltonian systems and gradient flows. The gradient structure is associated with conservation of invariants or with dissipation and hence plays a crucial role in the description of the physical properties of the system. Traditional hyper-reduction of nonlinear gradient fields yields efficient approximations that, however, lack the gradient structure. We focus on Hamiltonian gradients and we propose to first decompose the nonlinear part of the Hamiltonian, mapped into a suitable reduced space, into the sum of d terms, each characterized by a sparse dependence on the system state. Then, the hyper-reduced approximation is obtained via discrete empirical interpolation (DEIM) of the Jacobian of the derived d -valued nonlinear function. The resulting hyper-reduced model retains the gradient structure and its computational complexity is independent of the size of the full model. Moreover, *a priori* error estimates show that the hyper-reduced model converges to the reduced model and the Hamiltonian is asymptotically preserved. Whenever the evolution of the nonlinear Hamiltonian gradient requires high-dimensional DEIM approximation spaces, an adaptive strategy is performed. This consists in updating the hyper-reduced Hamiltonian via a low-rank correction of the DEIM basis. Numerical tests demonstrate the runtime speedups compared to the full and the reduced models.

Key words. adaptive hyper-reduction, discrete empirical interpolation, preservation of gradient structure, nonlinear Hamiltonian systems, symplectic model order reduction

AMS subject classifications. 65P10, 37N30, 65L99

1. Introduction. We consider parametric dynamical systems where the velocity describing the flow is characterized by nonlinear gradient fields, that is

$$(1.1) \quad \dot{y}(t, \eta) = X(y(t, \eta), \eta) \quad \text{for } t \geq t_0,$$

where $\eta \in \mathbb{R}^p$, $p \geq 1$, is a parameter, $y : (t_0, \infty) \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ is the state variable, and the velocity field $X : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ is assumed to be of the form $X(y, \eta) = S \nabla_y \mathcal{H}(y, \eta)$, with S in $\mathbb{R}^{n \times n}$ and \mathcal{H} a nonlinear function of the state. Examples of such dynamical systems are gradient flows, where S is a negative semi-definite matrix, Hamiltonian systems, where S is skew-symmetric, and dissipative Hamiltonian systems, such as port-Hamiltonian systems, where $S = J - R$ is the sum of a skew-symmetric component J and a symmetric positive semi-definite component R . The gradient structure of the velocity field X plays a crucial role in the characterization of the dynamics and of the physical properties of the system since it is associated with dissipation or preservation of quantities such as energy or entropy.

In this work we focus on Hamiltonian dynamical systems and the goal is to develop numerical approximation methods for their efficient solution in large-scale scenarios, i.e. when the number n of degrees of freedom is high, and for many instances of the parameter. A major challenge in this task is to preserve the physical properties of the dynamics and, in particular, the Hamiltonian gradient structure of the velocity field.

In the context of large-scale many-query simulations of parametric differential problems, model order reduction (MOR) and reduced basis methods (RBM) have been

*Submitted to the editors on June 2022.

[†]Centre for Analysis, Scientific computing and Applications, Eindhoven University of Technology (TU/e), The Netherlands. (c.pagliantini@tue.nl, f.vismara@tue.nl).

developed to provide efficient high-fidelity surrogate models on lower-dimensional spaces where the bulk of the dynamics takes place. Their success notwithstanding, the application of RBM to problems of the form (1.1) might lead to unstable and qualitatively wrong solution behavior since the gradient structure of the model can be destroyed during dimension reduction. To address this issue in solving Hamiltonian systems, structure-preserving MOR techniques have been developed in recent years to derive reduced models that retain (at least part of) the geometric structure underlying the dynamics. The first methods of this class were developed in [20] using a Lagrangian formulation and were later extended to nonlinear parametric Lagrangian systems in [6]; reduced models for canonical Hamiltonian systems on symplectic vector spaces were derived in [25] using POD-type strategies and in [1] via symplectic greedy algorithms; reduced bases that are symplectic but not orthogonal were proposed in [3, 4]; in [13] reduced models were obtained via separate reduced spaces for the generalized positions and momenta; in [29] a double projection of the original problem and of the Hamiltonian gradient into the span of a POD basis is employed; general Hamiltonian systems on Poisson manifolds were tackled in [17]; in [27] operator inference techniques are proposed to deal with linear Hamiltonian systems. We refer to [19] for a detailed overview on structure-preserving MOR for Hamiltonian systems. Although these methods have led to the successful construction of stable low-dimensional models, little attention has been paid to the efficient treatment of nonlinear operators. Indeed, in the presence of Hamiltonian functions with general nonlinear dependence on the state, the computational cost of solving the reduced model might still depend on the size of the underlying full model, resulting in simulation times that hardly improve over the original system simulation. This is a well-known issue in model order reduction and has led to the so-called hyper-reduction methods, which are usually based on approximations of the high-dimensional nonlinear operators using sparse sampling via interpolation among samples. Traditional hyper-reduction techniques, however, do not preserve gradient fields. This implies that hyper-reduced models of Hamiltonian systems are no longer Hamiltonian. As real-world processes tend to be nonlinear, a lack of efficient and physically compatible dimension reduction of general nonlinearities comes to the fore.

To the best of our knowledge, only a few works have considered hyper-reduction methods aimed at preserving Hamiltonian structures [6, 12, 8, 30, 18]. Although these techniques introduce interesting ideas, they appear tailored to specific problems or lack a rigorous theoretical analysis of the suggested approximations. In [6] the reduced Hamiltonian is approximated with a Taylor polynomial expansion truncated at the second term. As reported by the authors, the scheme is only effective for asymptotically stable systems and when the expansion is performed around an equilibrium point. These assumptions rule out many important cases and are generally not met by evolution problems stemming from the semi-discretization of partial differential equations. A cubature approach, the Energy-Conserving Sampling and Weighting (ECSW) scheme, is presented in [12]: the nonlinear vector field obtained from the semi-discretization of a Hamiltonian PDE is approximated with a weighted average of the field components on a coarser mesh. The method is, however, limited to finite element discretization of Hamiltonian PDEs, and requires a very expensive offline phase, especially for parametric problems. A variation of the discrete empirical interpolation method (DEIM) [2, 9] has been proposed in [8]: the nonlinear Hamiltonian gradient is approximated in the space where the DEIM projection is orthogonal. The Hamiltonian structure is preserved since orthogonal projections preserve gradients, but the method does not guarantee that the resulting approximation is accurate. A suitable

combination of DEIM and dynamic mode decomposition (DMD) is proposed in [18] to deal with the nonlinear operators of the Hamiltonian systems resulting from a particle-based discretization of the Vlasov–Poisson kinetic plasma model. In a recent preprint [30], DEIM hyper-reduction is applied to a nonlinear vector-valued function G obtained by decomposing the Hamiltonian into the Euclidean product of G with a constant vector. The gradient of the resulting operator provides an approximation of the original Hamiltonian gradient but there is no guarantee on the accuracy of such approximation nor on asymptotic convergence as the DEIM space is enlarged.

In this work we propose a novel gradient-preserving hyper-reduction strategy to construct surrogate models that: (i) are Hamiltonian; (ii) can be solved at a computational cost independent of the size of the full model; (iii) are provably accurate; and (iv) ensure preservation of the full order Hamiltonian asymptotically. The proposed approach targets the *reduced* nonlinear operator, in contrast to traditional hyper-reduction techniques where the nonlinear high-dimensional function is approximated first and then the resulting approximation is projected onto a reduced space. In particular, we first map the full order nonlinear Hamiltonian gradient into the reduced space via a structure-preserving (here symplectic) projection. Then, the nonlinear part of the reduced Hamiltonian is written as the sum of d terms, where d is typically of the order of the size of the full model. Although a similar decomposition is used in [30], here we do not approximate the resulting d -valued nonlinear function but rather the Hamiltonian gradient. This is done by projecting the Jacobian of the nonlinear function using discrete empirical interpolation (DEIM) [2, 9]. We derive *a priori* error estimates of the error between the full model solution and its hyper-reduced approximation and of the error in the conservation of the Hamiltonian. These results show that the hyper-reduced model is asymptotically as accurate as the reduced model. Since the target accuracy is the one of the reduced model, the hyper-reduction allows to obtain accurate solutions and to ensure exact Hamiltonian conservation with DEIM sizes much smaller than d and, hence, at a significantly reduced computational cost.

Whenever the dynamics of the problem does not allow for small DEIM approximations and, hence, enough computational savings, we let the proposed gradient-preserving hyper-reduction change in time. We derive an adaptive approach that extends to gradient fields the adaptive discrete empirical interpolation method (ADEIM) introduced in [24]. The DEIM basis is updated in time with the low-rank factor that minimizes a residual of the nonlinear Jacobian at suitably chosen sampling points. In particular, we extend the method of [24] to updates of general rank and to matrix-valued nonlinear terms. We also derive an algorithm for the optimal choice of the sampling points. A detailed analysis of the computational complexity of the adaptive hyper-reduction shows that, if the adaptation is implemented in an efficient way, the complexity reduction due to a smaller DEIM basis (compared to the non-adaptive algorithm) outweighs the extra computational cost of constructing the update.

The remainder of the paper is organized as follows. In [Section 2](#) we introduce Hamiltonian dynamical systems and discuss their symplectic model order reduction. [Section 3](#) concerns the structure-preserving DEIM hyper-reduction of the Hamiltonian gradient. [Section 4](#) is devoted to the adaptive gradient-preserving hyper-reduction method and a summary of the scheme is presented in [Algorithm 4.3](#). In [Section 5](#) the proposed methods are tested on a set of numerical experiments. Some conclusions and open questions are presented in [Section 6](#).

2. Full order system and its model order reduction. Let $\mathcal{T} := (t_0, T] \subset \mathbb{R}$ be a temporal interval, let $\mathcal{P} \subset \mathbb{R}^p$, with $p \geq 1$, be a compact set of parameters, and

let $V_{2n} \subset \mathbb{R}^{2n}$ be the phase space. For each $\eta \in \mathcal{P}$, we consider the Hamiltonian dynamical system: given $y^0(\eta) \in V_{2n}$, find $y(\cdot, \eta) \in C^1(\mathcal{T}; V_{2n})$ such that

$$(2.1) \quad \begin{cases} \dot{y}(t, \eta) = J_{2n} \nabla_y \mathcal{H}(y(t, \eta), \eta), & t \in \mathcal{T}, \\ y(t_0, \eta) = y^0(\eta), \end{cases}$$

where $y : \mathcal{T} \times \mathcal{P} \rightarrow V_{2n}$ is the state variable and the matrix $J_{2n} \in \mathbb{R}^{2n \times 2n}$, the so-called canonical Poisson tensor, is defined as

$$J_{2n} = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix}.$$

The function $\mathcal{H} : V_{2n} \times \mathcal{P} \rightarrow \mathbb{R}$ is the Hamiltonian of the system and we write it in the following general form

$$(2.2) \quad \mathcal{H}(y, \eta) = \frac{1}{2} y^\top L(\eta) y + y^\top f(\eta) + \mathcal{G}(\eta) + \mathcal{N}(y, \eta),$$

where $L(\eta) \in \mathbb{R}^{2n \times 2n}$ is a symmetric positive semi-definite matrix, $f(\eta) \in \mathbb{R}^n$, $\mathcal{G}(\eta) \in \mathbb{R}$ and $\mathcal{N}(\cdot, \eta)$ is a nonlinear function of the state variable. The Hamiltonian is a conserved quantity of system (2.1), namely it remains constant along solution trajectories. Moreover, the phase space of Hamiltonian systems has a symplectic geometric structure. Here we focus on symplectic vector spaces: this means that the space V_{2n} can be endowed with a local basis $\{e_i\}_{i=1}^{2n}$ which is symplectic and orthonormal [5, Chapter 12], that is $e_i^\top J_{2n} e_j = (J_{2n})_{i,j}$ and $(e_i, e_j) = \delta_{i,j}$ for all $i, j = 1 \dots, 2n$, where (\cdot, \cdot) is the Euclidean inner product. We refer to [21, Chapter 10] for a detailed introduction to Hamiltonian systems. To ensure well-posedness of (2.1) we assume that, for each parameter $\eta \in \mathcal{P}$, the Hamiltonian gradient $\nabla_y \mathcal{H}(\cdot, \eta)$ is Lipschitz continuous in the Euclidean norm uniformly with respect to time.

Throughout the paper, we use the symbols $\|\cdot\|$ and $\|\cdot\|_2$ to denote the vector and matrix 2-norm, respectively, while $\|\cdot\|_F$ refers to the Frobenius norm.

2.1. Symplectic model order reduction. When the number n of degrees of freedom in (2.1) is large and the system needs to be solved for many instances of the parameter η , model order reduction can be used to reduce the complexity of the original system and thus to speed up the resulting numerical simulations. Among model order reduction techniques, we consider reduced basis methods, which construct an approximation space of low dimension, the so-called *reduced space*, from a collection of simulation data corresponding to full order solutions at sampled values of time and parameters. A low-dimensional model is then obtained via a suitable projection of the full order dynamics onto the reduced space.

More in details, the state y is approximated as

$$y_{rb}(t, \eta) = \sum_{i=1}^{2k} z_i(t, \eta) a_i = A z(t, \eta), \quad \text{for all } t \in \mathcal{T}, \eta \in \mathcal{P},$$

where $k \ll n$, $A = [a_1 \dots a_{2k}] \in \mathbb{R}^{2n \times 2k}$ is the reduced basis and $z(t, \eta) \in \mathbb{R}^{2k}$ is the vector of expansion coefficients. In order to preserve the Hamiltonian structure of the full model, the matrix A is enforced to be symplectic, that is, $A^\top J_{2n} A = J_{2k}$, and orthogonal, that is, $A^\top A = I_{2k}$. If A is an orthogonal symplectic matrix, then $A^\top J_{2n} = J_{2k} A^\top$, see e.g. [25, Lemma 3.3]. Projecting the full order model (2.1) onto

the space V_{2k} spanned by the columns of A yields the reduced model: for any $\eta \in \mathcal{P}$, find $z(\cdot, \eta) \in C^1(\mathcal{T}; V_{2k})$ such that

$$(2.3) \quad \begin{cases} \dot{z}(t, \eta) = J_{2k} \nabla_z \mathcal{H}_r(z(t, \eta), \eta), & t \in \mathcal{T}, \\ z(t_0, \eta) = z^0(\eta) := A^\top y^0(\eta), \end{cases}$$

where the reduced Hamiltonian is obtained from (2.2) as

$$\mathcal{H}_r(z, \eta) := \mathcal{H}(Az, \eta) = \frac{1}{2} z^\top L_r(\eta) z + z^\top f_r(\eta) + \mathcal{G}(\eta) + \mathcal{N}_r(z, \eta),$$

with $L_r(\eta) := A^\top L(\eta) A \in \mathbb{R}^{2k \times 2k}$, $f_r(\eta) := A^\top f(\eta) \in \mathbb{R}^{2k}$ and $\mathcal{N}_r := \mathcal{N} \circ A$.

An orthogonal and symplectic reduced basis can be constructed from a set of full model solutions using SVD-type strategies, such as cotangent lift, complex SVD, or nonlinear programming [25], or with the symplectic greedy algorithm of [1].

3. Hyper-reduction of the Hamiltonian gradient. Assuming the affine separability of the operators $L(\eta)$ and $f(\eta)$ in (2.2), the major computational cost in solving the reduced model (2.3) comes from the nonlinear term \mathcal{N}_r of the Hamiltonian. If standard hyper-reduction techniques are applied to the reduced Hamiltonian gradient $\nabla_z \mathcal{H}_r$, the resulting approximate function is generally no longer a gradient field, which might lead to unstable or inaccurate approximations. Numerical evidence of this behavior is reported e.g. in [25]. To address this shortcoming we first consider a decomposition of the nonlinear part of the reduced Hamiltonian as

$$(3.1) \quad \mathcal{N}_r(z, \eta) = \sum_{i=1}^d v_i h_i(Az, \eta) = v^\top h(Az, \eta), \quad \text{for all } z \in V_{2k}, \eta \in \mathcal{P},$$

where $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ is a constant vector and $h : V_{2n} \times \mathcal{P} \rightarrow \mathbb{R}^d$. A decomposition of the Hamiltonian of the form (3.1) has appeared also in [30] and holds in many cases of interest. For example, the Hamiltonian given by the energy of a system of particles can be written as the sum of the contribution of each particle; or, if (2.1) results from the semi-discretization of a PDE via a local approximation, e.g. finite element or finite volume schemes, h_i represents the contribution of the i th mesh element to the total Hamiltonian. We refer to [21] and [15, Chapter I] for several examples of Hamiltonian in this form and we will further comment on this aspect in Remark 3.1.

Using the decomposition (3.1), the gradient of the nonlinear term \mathcal{N}_r reads

$$(3.2) \quad \nabla_z \mathcal{N}_r(z, \eta) = A^\top J_{\eta, h}^\top(Az) v, \quad \text{for all } z \in V_{2k}, \eta \in \mathcal{P},$$

where $J_{\eta, h}(\cdot) \in \mathbb{R}^{d \times 2n}$ is the Jacobian of $h(\cdot, \eta)$. The idea is then to use discrete empirical interpolation to approximate the term $J_{\eta, h}(A \cdot) A \in \mathbb{R}^{d \times 2k}$, namely the Jacobian mapped to the reduced space. The projected reduced Jacobian is

$$(3.3) \quad \mathbb{P} J_{\eta, h}(A \cdot) A, \quad \text{where} \quad \mathbb{P} := U(P^\top U)^{-1} P^\top \in \mathbb{R}^{d \times d}.$$

Here, $U \in \mathbb{R}^{d \times m}$ is the so-called DEIM basis with $m \ll d$, and $P := [\mathbf{e}_{\beta(1)}, \dots, \mathbf{e}_{\beta(m)}]$, where \mathbf{e}_i is the i th unit vector of \mathbb{R}^d and $\{\beta(1), \dots, \beta(m)\} \subset \{1, \dots, d\}$ are interpolation indices [9]. The DEIM projection introduced in (3.3) can be equivalently seen as an approximation of the nonlinear term \mathcal{N}_r in (3.1) by

$$(3.4) \quad \mathcal{N}_{hr}(z, \eta) := v^\top \mathbb{P} h(Az, \eta) \quad \text{for all } z \in V_{2k}, \eta \in \mathcal{P}.$$

With the nonlinear term of the Hamiltonian approximated as in (3.4), the resulting hyper-reduced model reads: for any $\eta \in \mathcal{P}$, find $z(\cdot, \eta) \in C^1(\mathcal{T}; V_{2k})$ such that

$$(3.5) \quad \begin{cases} \dot{z}(t, \eta) = J_{2k} \nabla_z \mathcal{H}_{hr}(z(t, \eta), \eta), & t \in \mathcal{T}, \\ z(t_0, \eta) = z^0(\eta) := A^\top y^0(\eta), \end{cases}$$

The dynamical system (3.5) is still Hamiltonian, with the hyper-reduced Hamiltonian defined as

$$\mathcal{H}_{hr}(z, \eta) = \frac{1}{2} z^\top L_r(\eta) z + z^\top f_r(\eta) + \mathcal{G}(\eta) + \mathcal{N}_{hr}(z, \eta),$$

and its gradient is given by $\nabla_z \mathcal{H}_{hr}(z, \eta) = L_r(\eta) z + f_r(\eta) + A^\top J_{\eta, h}^\top(Az) \mathbb{P}^\top v$.

Remark 3.1. A key role in this approach is played by the decomposition (3.1), which is not unique, and by the choice of d . Indeed, let us assume that, for fixed $z \in V_{2k}$ and $\eta \in \mathcal{P}$, each entry of $h(y, \eta)$, with $y := Az$, depends on $s_1 \leq n$ components among the first n entries of y and on $s_2 \leq n$ components among the last n entries of y . Then, every row of $J_{\eta, h}$ has at most $s_1 + s_2$ non-zero elements. This implies that the evaluation of the gradient (3.2) of the reduced Hamiltonian has complexity $O((s_1 + s_2)dk)$, while the cost of computing the gradient of the hyper-reduced Hamiltonian (3.4) is of order $O((s_1 + s_2)mk)$. Based on this observation, one might be tempted to choose a decomposition (3.1) with a small d and bypass the hyper-reduction entirely. This choice is however often associated with large values of s_1 and s_2 since typically $(s_1 + s_2)d$ is, at least, of the order of n . The choice of the decomposition should instead be driven as to maximize the sparsity of the Jacobian matrix $J_{\eta, h} \in \mathbb{R}^{d \times 2n}$, hence maximizing the efficacy of the hyper-reduction. We refer to the numerical experiments in Section 5 for concrete examples of such decomposition.

3.1. Construction of the DEIM projection. As anticipated in (3.3), we propose an approximation that minimizes the DEIM projection error of the reduced Jacobian, $J_{\eta, h}(A \cdot) A \in \mathbb{R}^{d \times 2k}$, once the reduced basis A has been constructed. The algorithm of [30] seeks instead to approximate the factor h in (3.1), but this comes with no guarantee on the accuracy of the approximation of the reduced Hamiltonian gradient and, in turn, on the quality of the hyper-reduced model (3.5).

To build the DEIM projection, we compute snapshots in the reduced space V_{2k} as $\{A^\top y^{\ell_i}(\eta_j)\}_{i,j}$ for $j = 1, \dots, N_p$ and $i = 1, \dots, N_s$, where $y^{\ell_i}(\eta_j)$ is the full model solution at time t^{ℓ_i} and parameter η_j . We collect the corresponding snapshots of the reduced Jacobian in the matrix

$$(3.6) \quad M_J = [J_{\eta_1, h}(AA^\top y^{\ell_1}(\eta_1))A \dots J_{\eta_{N_p}, h}(AA^\top y^{\ell_{N_s}}(\eta_{N_p}))A] \in \mathbb{R}^{d \times 2k N_s N_p}.$$

The DEIM basis $U \in \mathbb{R}^{d \times m}$ in (3.3) is obtained from M_J via proper orthogonal decomposition (POD). Observe that the snapshot matrix M_J can have a potentially large number of columns: to curb the high computational cost to derive the DEIM basis from M_J , one could make use of randomized algorithms [16].

Once the DEIM basis is fixed, there are several ways to select the DEIM interpolation indices so that $P^\top U$ is non-singular. In this work we focus on the greedy algorithm of [9, Algorithm 1]. An alternative choice is, for example, QDEIM [11] which is based on a QR factorization with column pivoting of the matrix U .

Remark 3.2. In principle, one may approximate the full Jacobian $J_{\eta, h}$ rather than the reduced Jacobian $J_{\eta, h}(A \cdot) A$. However, there are (at least) two major issues in

dealing with the full Jacobian. First, the snapshot matrix M_J has $2nN_sN_p$ columns instead of $2kN_sN_p$, where $n \gg k$. In many applications, already for one-dimensional problems, this matrix is too large to be stored, and the cost of computing its singular values and vectors is prohibitive. This is particularly relevant in the adaptive approach, see [Section 4](#), when the DEIM projection is updated online. Second, the full Jacobian $J_{\eta,h}$ usually exhibits diagonal (or sparsity) patterns so that the associated snapshot matrix has (almost) full rank. This implies that one has to select $m \approx d$ DEIM basis functions to achieve a sufficiently accurate approximation, making hyper-reduction basically ineffective. On the contrary, when the reduced Jacobian is considered, it is found in practice that the singular values of M_J exhibit a much faster decay, which allows for an efficient hyper-reduction. We refer to [Figure 4](#) in [Subsection 5.2](#) for a numerical example that shows this behavior.

3.2. A priori convergence estimates. The error between the full order and hyper-reduced solution can be bounded by the error between the full order solution and its projection onto the reduced space and by the DEIM projection error of the reduced Jacobian. The following theorem extends to our setting the result in [\[10, Theorem 3.1\]](#).

THEOREM 3.3. *For any $\eta \in \mathcal{P}$, let $y(\cdot, \eta) \in C^1(\mathcal{T}, V_{2n})$ be the solution of the full model [\(2.1\)](#) and let $z(\cdot, \eta) \in C^1(\mathcal{T}, V_{2k})$ be the solution of the hyper-reduced system [\(3.5\)](#). Let \mathbb{P} be the DEIM projection [\(3.3\)](#). Assume that, for every $\eta \in \mathcal{P}$, the Jacobian $J_{\eta,h}$ of h [\(3.1\)](#) is Lipschitz continuous in the 2-norm with constant $\mathcal{L}_h(\eta)$. Then,*

$$(3.7) \quad \begin{aligned} \|y - Az\|_{L^2(\mathcal{T} \times \mathcal{P}; V_{2n})}^2 &\leq C_1(T) \|y - AA^\top y\|_{L^2(\mathcal{T} \times \mathcal{P}; V_{2n})}^2 \\ &\quad + C_2(T) \|(I - \mathbb{P})J_{\eta,h}(AA^\top y)A\|_{L^2(\mathcal{T} \times \mathcal{P}; V_{2n})}^2, \end{aligned}$$

where $C_1(T) := \Delta\mathcal{T} \max_{\eta \in \mathcal{P}} (C_\eta(T)\alpha^2(\eta)) + 1$ and $C_2(T) := \Delta\mathcal{T} \|v\|^2 \max_{\eta \in \mathcal{P}} C_\eta(T)$,

$$C_\eta(T) := \begin{cases} \beta^{-1}(\eta)(e^{2\beta(\eta)(T-t_0)} - 1), & \text{if } \beta(\eta) \neq 0, \\ 2(T - t_0), & \text{if } \beta(\eta) = 0, \end{cases}$$

with $\alpha(\eta) := \|A^\top K(\eta)\|_2 + \mathcal{L}_\mathcal{N}(\eta)$, $\mathcal{L}_\mathcal{N}(\eta)$ Lipschitz continuity constant of $\nabla\mathcal{N}(\cdot, \eta)$, $\beta(\eta) := \mu(K(\eta)) + \mathcal{L}_h(\eta) \|(P^\top U)^{-1}\|_2 \|v\|$, and $\mu(K(\eta))$ logarithmic norm of $K(\eta) := J_{2n}L(\eta)$ with respect to the 2-norm, i.e., $\mu(K) := \sup_{x \neq 0} \frac{(x, Kx)}{(x, x)}$.

Proof. Let us fix the parameter $\eta \in \mathcal{P}$. The error at each $t \in \mathcal{T}$ can be split as

$$e(t, \eta) := y(t, \eta) - Az(t, \eta) = e_p(t, \eta) + e_h(t, \eta),$$

where $e_p := y - AA^\top y$ is the projection error and $e_h := AA^\top y - Az$. Differentiating e_h with respect to time, using [\(2.1\)](#) and [\(3.5\)](#), and the symplecticity of A results in

$$\dot{e}_h(t, \eta) = AA^\top \dot{y}(t, \eta) - A\dot{z}(t, \eta) = AA^\top K(\eta)e_h(t, \eta) + s(t, \eta),$$

where $s(t, \eta) := AA^\top K(\eta)e_p(t, \eta) + AJ_{2k}(A^\top \nabla_y \mathcal{N}(y, \eta) - \nabla_z \mathcal{N}_{hr}(z, \eta))$ and $K(\eta) := J_{2n}L(\eta)$. The norm of s can be bounded as

$$(3.8) \quad \|s(t, \eta)\| \leq \|A^\top K(\eta)\|_2 \|e_p(t, \eta)\| + \|A^\top \nabla_y \mathcal{N}(y, \eta) - \nabla_z \mathcal{N}_{hr}(z, \eta)\|.$$

The second term in (3.8) can be split as follows,

$$(3.9) \quad \begin{aligned} \|A^\top \nabla_y \mathcal{N}(y, \eta) - \nabla_z \mathcal{N}_{hr}(z, \eta)\| &\leq \|A^\top \nabla_y \mathcal{N}(y, \eta) - A^\top \nabla_y \mathcal{N}(AA^\top y, \eta)\| \\ &\quad + \|A^\top \nabla_y \mathcal{N}(AA^\top y, \eta) - \nabla_z \mathcal{N}_{hr}(A^\top y, \eta)\| \\ &\quad + \|\nabla_z \mathcal{N}_{hr}(A^\top y, \eta) - \nabla_z \mathcal{N}_{hr}(z, \eta)\|. \end{aligned}$$

We consider the three terms on the right-hand side separately. First observe that, if $\nabla_y \mathcal{H}(\cdot, \eta)$ is Lipschitz continuous in the 2-norm with constant $\mathcal{L}_{\mathcal{H}}(\eta)$, then $\nabla \mathcal{N}(\cdot, \eta)$ is also Lipschitz continuous with constant $\mathcal{L}_{\mathcal{N}}(\eta) \leq \mathcal{L}_{\mathcal{H}}(\eta) + \|L(\eta)\|_2$. This gives

$$\|A^\top \nabla_y \mathcal{N}(y, \eta) - A^\top \nabla_y \mathcal{N}(AA^\top y, \eta)\| \leq L_{\mathcal{N}}(\eta) \|e_p(t, \eta)\|.$$

Moreover, using the decomposition of the Hamiltonian in (3.1) results in

$$\begin{aligned} \|A^\top \nabla_y \mathcal{N}(AA^\top y, \eta) - \nabla_z \mathcal{N}_{hr}(A^\top y, \eta)\| &= \|A^\top J_{\eta, h}^\top(AA^\top y)v - A^\top J_{\eta, h}^\top(AA^\top y)\mathbb{P}^\top v\| \\ &\leq \|v\| \|(I - \mathbb{P})J_{\eta, h}(AA^\top y)A\|_2, \end{aligned}$$

and similarly

$$(3.10) \quad \begin{aligned} \|\nabla_z \mathcal{N}_{hr}(A^\top y, \eta) - \nabla_z \mathcal{N}_{hr}(z, \eta)\| &= \|A^\top J_{\eta, h}^\top(AA^\top y)\mathbb{P}^\top v - A^\top J_{\eta, h}^\top(Az)\mathbb{P}^\top v\| \\ &\leq L_h(\eta) \|AA^\top y - Az\| \|\mathbb{P}^\top v\| \leq \gamma(\eta) \|e_h(t, \eta)\|, \end{aligned}$$

where $\gamma(\eta) := L_h(\eta) \|(P^\top U)^{-1}\|_2 \|v\|$. We can now substitute these three bounds into (3.9) and (3.8) to obtain

$$(3.11) \quad \|s(t, \eta)\| \leq \alpha(\eta) \|e_p(t, \eta)\| + \gamma(\eta) \|e_h(t, \eta)\| + \|v\| \|W(t, \eta)\|_2,$$

where $\alpha(\eta) := \|A^\top K(\eta)\|_2 + L_{\mathcal{N}}(\eta)$, and $W(t, \eta) := (I - \mathbb{P})J_{\eta, h}(AA^\top y(t, \eta))A$.

The time derivative of the error norm gives

$$(3.12) \quad \frac{d}{dt} \|e_h(t, \eta)\| = \frac{1}{\|e_h(t, \eta)\|} \left((AA^\top K(\eta)e_h(t, \eta), e_h(t, \eta)) + (s(t, \eta), e_h(t, \eta)) \right).$$

The first term on the right-hand side of (3.12) can be bounded as in [10, Theorem 3.1], by $\mu(K(\eta)) \|e_h(t, \eta)\|$, using the fact that $(AA^\top K(\eta)e_h, e_h) = (AA^\top K(\eta)A\tilde{e}_h, A\tilde{e}_h) \leq \mu(A^\top K A) \|\tilde{e}_h\|^2$, with $\tilde{e}_h := A^\top y - z$.

Using (3.11) for the second term of (3.12) yields

$$\frac{d}{dt} \|e_h(t, \eta)\| \leq \beta(\eta) \|e_h(t, \eta)\| + b(t, \eta),$$

where $\beta(\eta) := \mu(K(\eta)) + \gamma(\eta)$, and $b(t, \eta) := \alpha(\eta) \|e_p(t, \eta)\| + \|v\| \|W(t, \eta)\|_2$.

A standard application of the Gronwall inequality [14] yields the conclusion. \square

The dependence of the constants in the bound (3.7) on the norm of the vector v can be avoided by a suitable normalization of the decomposition (3.1). Observe that the first term on the right-hand side of (3.7) is controlled via the reduced space, while the second term can be minimized in the construction of the DEIM basis. Indeed, the error of the approximation of the reduced Jacobian given by the DEIM projection (3.3) can be bounded analogously to [9, Lemma 3.2] as

$$\|(I - \mathbb{P})J_{\eta_j, h}(AA^\top y^{s_i}(\eta_j))A\|_2 \leq \|(P^\top U)^{-1}\|_2 \sqrt{\sum_{\ell=m+1}^{\text{rank}(M_J)} \sigma_\ell^2},$$

for any $i = 1, \dots, N_s$, and $j = 1, \dots, N_p$, where $\{\sigma_\ell\}_\ell$ are the singular values of M_J .

For Jacobian matrices having a particular structure, the Lipschitz continuity assumption of $J_{\eta,h}$ in [Theorem 3.3](#) might not be needed. An example often encountered in applications is when the i th entry of $h(y, \eta)$ only depends on the i th pair of symplectic variables, namely on the i th and $(i+n)$ th components of y , for any $y \in V_{2n}$. In this case, $d = n$ and $J_{\eta,h} \in \mathbb{R}^{n \times 2n}$ is formed of two $n \times n$ diagonal blocks. This implies that $\nabla_z \mathcal{N}_{hr} = J_{\eta,h}^\top \mathbb{P}^\top v$ is of the form $D \nabla \mathcal{N}$ where

$$D := \text{diag}((\mathbb{P}^\top v)_1, \dots, (\mathbb{P}^\top v)_n, (\mathbb{P}^\top v)_1, \dots, (\mathbb{P}^\top v)_n) \in \mathbb{R}^{2n \times 2n}.$$

Therefore, the bound in [\(3.10\)](#) becomes

$$\begin{aligned} \|\nabla_z \mathcal{N}_{hr}(A^\top y, \eta) - \nabla_z \mathcal{N}_{hr}(z, \eta)\| &= \|A^\top D(\nabla_y \mathcal{N}(AA^\top y, \eta) - \nabla_y \mathcal{N}(Az, \eta))\| \\ &\leq \|D\|_2 \mathcal{L}_{\mathcal{N}}(\eta) \|AA^\top y - Az\|, \end{aligned}$$

where $\|D\|_2 \leq \|\mathbb{P}^\top v\|_2 \leq \|(P^\top U)^{-1}\|_2 \|v\|$. Hence, [\(3.10\)](#) holds with the constant $\gamma(\eta) := L_{\mathcal{N}}(\eta) \|(P^\top U)^{-1}\|_2 \|v\|$ and no Lipschitz continuity of $J_{\eta,h}$ is required.

3.3. Conservation of the Hamiltonian. In this section we assess the error in the conservation of the Hamiltonian introduced by the hyper-reduction. In the following result we show that the error between the Hamiltonian evaluated at the full model solution and at the hyper-reduced solution is bounded by the approximation error of \mathcal{N} . This implies that the hyper-reduced model guarantees preservation of the Hamiltonian for sufficiently large DEIM bases, that is, when the accuracy of the hyper-reduced model reaches the accuracy of the reduced model and up to errors due to the time integration scheme. We refer to [Section 5](#) for several numerical tests that corroborate this result.

PROPOSITION 3.4. *Let $\eta \in \mathcal{P}$ be fixed. Let $y^j(\eta)$ be an approximation of the solution $y(t^j, \eta)$ of the full order system [\(2.1\)](#) at time t^j , with $j = 1, \dots, N_t$, obtained with a user-defined numerical time integrator. Similarly, let $z^j(\eta)$ be an approximation of the solution $z(t^j, \eta)$ of the hyper-reduced system [\(3.5\)](#) at time t^j . Assume that $y^0(\eta) \in \text{Col}(A)$ and $h(AA^\top y^0(\eta), \eta) \in \text{Col}(U)$, where Col denotes the column space. Then, the error $\Delta \mathcal{H}_j(\eta) := |\mathcal{H}(y^j(\eta), \eta) - \mathcal{H}(Az^j(\eta), \eta)|$ satisfies*

$$\Delta \mathcal{H}_j(\eta) \leq |v^\top (I - \mathbb{P})h(Az^j(\eta), \eta)| + \varepsilon_{\mathcal{H}}^{[t^0, t^j]} + \varepsilon_{\mathcal{H}_{hr}}^{[t^0, t^j]},$$

where $\varepsilon_{\mathcal{H}}^{[t^0, t^j]} := |\mathcal{H}(y^j) - \mathcal{H}(y^0)|$ and $\varepsilon_{\mathcal{H}_{hr}}^{[t^0, t^j]} := |\mathcal{H}_{hr}(z^j) - \mathcal{H}_{hr}(z^0)|$ are the errors in the Hamiltonian conservation, in the temporal interval $[t^0, t^j]$, associated with the chosen temporal integrator.

Proof. In this proof we omit the dependence of y^j , z^j and \mathcal{H} on η . The Hamiltonian error at the generic time instant t^j can be bounded as

$$(3.13) \quad \begin{aligned} \Delta \mathcal{H}_j(\eta) &\leq |\mathcal{H}(y^j) - \mathcal{H}(y^0)| + |\mathcal{H}_{hr}(z^j) - \mathcal{H}_{hr}(z^0)| \\ &\quad + |\mathcal{H}(Az^0) - \mathcal{H}_{hr}(z^0)| + |\mathcal{H}(Az^j) - \mathcal{H}_{hr}(z^j)| + \Delta \mathcal{H}_0(\eta). \end{aligned}$$

The first two terms in [\(3.13\)](#) measure the Hamiltonian conservation with respect to the initial value in the full and hyper-reduced systems. These quantities only depend on the time integration scheme. The term $|\mathcal{H}(Az^0) - \mathcal{H}_{hr}(z^0)| = |v^\top (I - \mathbb{P})h(Az^0)|$ vanishes since $h(Az^0) \in \text{Col}(U)$ by assumption. Similarly, the term $\Delta \mathcal{H}_0(\eta)$ vanishes due to the injectivity of the linear map A and the assumption $y^0 \in \text{Col}(A)$.

The conclusion follows by the definition of the full and reduced Hamiltonian. \square

The bound on the Hamiltonian error hinges on the assumptions that $y^0(\eta) \in \text{Col}(A)$ and $h(AA^\top y^0, \eta) \in \text{Col}(U)$ for any value of the parameter $\eta \in \mathcal{P}$. These conditions can be enforced via a shifting of the state variable and of the operator h . Let us introduce the variable $y_s(t, \eta) := y(t, \eta) - y^0(\eta)$ and the shifted operator $h_s(y(t, \eta), \eta) := h(y(t, \eta), \eta) - h(y^0(\eta), \eta)$, for any $t \in \mathcal{T}$ and $\eta \in \mathcal{P}$. Substituting into the expression of the Hamiltonian (2.2) gives

$$(3.14) \quad \mathcal{H}_s(y_s, \eta) := \mathcal{H}(y_s + y^0, \eta) = \frac{1}{2} y_s^\top L(\eta) y_s + y_s^\top f_s(\eta) + \mathcal{G}_s(\eta) + v^\top h_s(y_s + y^0, \eta),$$

where $f_s(\eta) := f(\eta) + L(\eta)y^0$ and $\mathcal{G}_s(\eta) = \mathcal{G}(\eta) + 1/2(y^0)^\top L(\eta)y^0 + (y^0)^\top f(\eta) + v^\top h(y^0, \eta)$. Then, the shifted state y_s satisfies an Hamiltonian system with Hamiltonian \mathcal{H}_s as in (3.14) and initial condition $y_s(t_0, \eta) = 0$. Since $y_s^0(\eta) = 0$ and $h_s(AA^\top y_s^0 + y^0, \eta) = h_s(y^0, \eta) = h(y^0, \eta) - h(y^0, \eta) = 0$ for any $\eta \in \mathcal{P}$, the shifted system satisfies the hypotheses of Proposition 3.4. The shifted system is the one we consider as full order model in all numerical tests of Section 5.

4. Adaptive gradient-preserving hyper-reduction. In many cases of interest, such as convection-dominated phenomena and conservative dynamics, it is known that the solution space, under variation of time and parameter, has poor global reducibility properties. An analogous property is observed when considering the space of Hamiltonian gradients, although a rigorous connection between the reducibility of the two spaces is not known. We refer to Subsection 5.2 for numerical evidence of these considerations.

To address this lack of global reducibility, we derive an adaptive strategy for the hyper-reduction of (2.1) where the DEIM projection is updated in time while preserving the gradient structure of the Hamiltonian vector field. The proposed approach is inspired by the method introduced in [24, 23].

For the sake of exposition, we assume that basis and interpolation points are adapted every δ time instants starting from t^{δ_0} , where $\delta, \delta_0 > 1$ are fixed hyper-parameters. Assuming that $N_t - \delta_0$ is a multiple of δ , the number of adaptations is $N_a = \delta^{-1}(N_t - \delta_0)$ and the adaptations are performed at time instants $\{t^{\ell_j}\}_{j=1}^{N_a}$, where $\ell_j := \delta_0 + (j-1)\delta$. We define $\ell_0 := 0$ and note that $t^{\ell_{N_a+1}} = t^{N_t} = T$. In principle, N_a does not need to be fixed *a priori* and it may be determined during time evolution of the hyper-reduced system based on e.g. suitable error criteria. In the proposed adaptive hyper-reduction algorithm, the basis update is performed for each instance of the test parameter. Let us then fix the problem parameter $\eta \in \mathcal{P}$.

Between two updates the local hyper-reduced system to be solved in the temporal interval $[t^{\ell_j}, t^{\ell_{j+1}}]$ reads

$$(4.1) \quad \begin{cases} \dot{z}(t, \eta) = J_{2k} \nabla_z \mathcal{H}_{hr}^j(z(t, \eta), \eta), & t \in (t^{\ell_j}, t^{\ell_{j+1}}], \\ z(t^{\ell_j}, \eta) = z^{\ell_j}(\eta), \end{cases}$$

for $j = 0, \dots, N_a$, where the local hyper-reduced Hamiltonian is given by

$$\mathcal{H}_{hr}^j(z, \eta) := \frac{1}{2} z^\top L_r(\eta) z + z^\top f_r(\eta) + \mathcal{G}(\eta) + v^\top \mathbb{P}_j h(Az, \eta),$$

and $\mathbb{P}_j := U_j(P_j^\top U_j)^{-1} P_j^\top$ is the local DEIM projection.

Starting from a DEIM basis $U_0 \in \mathbb{R}^{d \times m}$ and interpolation matrix $P_0 \in \mathbb{R}^{d \times m}$, the idea of the adaptive hyper-reduction is to perform a rank- r update of the DEIM basis to adapt the DEIM pair (U_j, P_j) to (U_{j+1}, P_{j+1}) for $j = 0, \dots, N_a - 1$. The initial

DEIM basis U_0 is constructed in the offline phase from snapshots of the full model solutions in the first δ_0 time steps, namely

$$M_J = [J_{\eta,h}(AA^\top y^1(\eta))A \dots J_{\eta,h}(AA^\top y^{\delta_0}(\eta))A] \in \mathbb{R}^{d \times 2k\delta_0}.$$

Observe that one might also construct the initial DEIM pair (U_0, P_0) from a set of training parameters.

Next, at time instant $t^{\ell_{j+1}}$, with $j = 0, \dots, N_a - 1$, the DEIM pair (U_j, P_j) is updated based on the snapshots of the nonlinear Hamiltonian in a past temporal window of size $w \in \mathbb{N}$, with $w < \delta_0$. More in details, consider the snapshot matrix of the reduced Jacobian in the temporal window $[t^{\ell_{j+1}-w+1}, t^{\ell_{j+1}}]$, namely let

$$F_j := [J_{\eta,h}(Az^{\ell_{j+1}-w+1})A \dots J_{\eta,h}(Az^{\ell_{j+1}})A] \in \mathbb{R}^{d \times \bar{w}},$$

where $\bar{w} := 2kw$. The residual of the DEIM approximation of the columns of F_j is given by $R_j = U_j C_j - F_j \in \mathbb{R}^{d \times \bar{w}}$, where $C_j \in \mathbb{R}^{m \times \bar{w}}$ is the DEIM coefficient matrix $C_j := (P_j^\top U_j)^{-1} P_j^\top F_j$. Analogously to [24], we update the DEIM basis matrix U_j to U_{j+1} via a rank- r correction, that is

$$(4.2) \quad U_{j+1} = U_j + \alpha_j \mathbf{b}_j^\top$$

with $\alpha_j \in \mathbb{R}^{d \times r}$ and $\mathbf{b}_j \in \mathbb{R}^{m \times r}$ of rank $r \in \mathbb{N}$, $r \leq m$. The update $\alpha_j \mathbf{b}_j^\top$ is defined as the rank- r matrix that minimizes the Frobenius norm of the residual at the sampling points collected in the matrix $S_j = [\mathbf{e}_{s_1^{(j)}} \dots \mathbf{e}_{s_{m_s}^{(j)}}] \in \mathbb{R}^{d \times m_s}$ where $s_1^{(j)}, \dots, s_{m_s}^{(j)} \in \{1, \dots, d\}$, namely it minimizes

$$\|S_j^\top (U_{j+1} C_j - F_j)\|_F^2 = \|S_j^\top R_j + S_j^\top \alpha_j \mathbf{b}_j^\top C_j\|_F^2.$$

With the change of variable $\mathbf{a}_j := S_j^\top \alpha_j \in \mathbb{R}^{m_s \times r}$, the update (4.2) boils down to solving the minimization problem

$$(4.3) \quad (\mathbf{a}_j, \mathbf{b}_j) = \arg \min_{(\mathbf{a}, \mathbf{b}) \in \mathcal{V}_r(m_s, m)} \|S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j\|_F^2,$$

over the space of rank- r matrices defined as

$$(4.4) \quad \mathcal{V}_r(m_s, m) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{m_s \times r} \times \mathbb{R}^{m \times r} : \text{rank}(\mathbf{a}) = \text{rank}(\mathbf{b}) = r\}.$$

The number m_s of sampling points is taken such that $m_s \geq m$ and $m_s \ll d$. The first requirement ensures that the minimization problem (4.3) has a non-trivial solution, while the second condition is enforced to avoid working in the potentially high dimension d . Notice that, since only m_s of the d rows of $S_j \mathbf{a}_j \in \mathbb{R}^{d \times r}$ are non-zero, the update (4.2) only modifies m_s rows of the DEIM basis U_j , that is

$$(4.5) \quad \check{S}_j^\top U_{j+1} = \check{S}_j^\top U_j \quad \text{where} \quad \check{S}_j := [\mathbf{e}_{s_{m_s+1}^{(j)}} \dots \mathbf{e}_{s_d^{(j)}}] \in \mathbb{R}^{d \times (d-m_s)},$$

and $\{s_{m_s+1}^{(j)}, \dots, s_d^{(j)}\} := \{1, \dots, d\} \setminus \{s_1^{(j)}, \dots, s_{m_s}^{(j)}\}$.

4.1. Rank- r update of the DEIM basis. The minimization problem (4.3) has been already considered in [24, Lemma 3.5] for rank-1 updates of the DEIM basis. We extend the result to the general case of rank $r > 1$.

THEOREM 4.1. *Let $r \leq m$ and assume that $C_j \in \mathbb{R}^{m \times \overline{m}}$ has full row-rank m . Let $\mathcal{V}_r(m_s, m)$ be defined as in (4.4). Then, the solution of the minimization problem*

$$(4.6) \quad \min_{(\mathbf{a}, \mathbf{b}) \in \mathcal{V}_r(m_s, m)} \left\| S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j \right\|_F^2,$$

is given by $\mathbf{a} = [a_1 \dots a_r] \in \mathbb{R}^{m_s \times r}$ and $\mathbf{b} = [b_1 \dots b_r] \in \mathbb{R}^{m \times r}$ where b_1, \dots, b_r are the eigenvectors of the generalized eigenvalue problem

$$(4.7) \quad C_j (S_j^\top R_j)^\top (S_j^\top R_j) C_j^\top v = \lambda C_j C_j^\top v,$$

corresponding to the r largest eigenvalues $\lambda_1 \geq \dots \geq \lambda_r \geq 0$, and

$$(4.8) \quad a_i = -\frac{1}{\|C_j^\top b_i\|^2} S_j^\top R_j C_j^\top b_i \quad i = 1, \dots, r.$$

Moreover, the solution $(\mathbf{a}, \mathbf{b}) \in \mathcal{V}_r(m_s, m)$ of (4.6) satisfies

$$(4.9) \quad \left\| S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j \right\|_F^2 = \left\| S_j^\top R_j \right\|_F^2 - \sum_{i=1}^r \lambda_i.$$

Proof. Let \mathcal{O}_j denote the objective function of the minimization problem (4.6), namely $\mathcal{O}_j(\mathbf{a}, \mathbf{b}) := \left\| S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j \right\|_F^2$. We recast problem (4.6) in the space

$$\widehat{\mathcal{V}}_r(m_s, m) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{m_s \times r} \times \mathbb{R}^{m \times r} : \text{rank}(\mathbf{a}) = \text{rank}(\mathbf{b}) = r, \mathbf{a}^\top \mathbf{a} \text{ diagonal}\}.$$

Assume that the pair $(\mathbf{c}, \mathbf{d}) \in \mathcal{V}_r(m_s, m)$ minimizes (4.6). Let $\mathbf{c} = \mathbf{f}Z$ be the thin QR factorization of \mathbf{c} without normalization, where $\mathbf{f} \in \mathbb{R}^{m_s \times r}$ and $Z \in \mathbb{R}^{r \times r}$ is upper triangular with ones on the main diagonal. Define $\mathbf{g} := \mathbf{d}Z^\top \in \mathbb{R}^{m \times r}$. Then $(\mathbf{f}, \mathbf{g}) \in \widehat{\mathcal{V}}_r(m_s, m)$ and $\mathbf{f} \mathbf{g}^\top C_j = \mathbf{c} \mathbf{d}^\top C_j$, which implies

$$\min_{(\mathbf{a}, \mathbf{b}) \in \mathcal{V}_r(m_s, m)} \left\| S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j \right\|_F^2 = \min_{(\mathbf{a}, \mathbf{b}) \in \widehat{\mathcal{V}}_r(m_s, m)} \left\| S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j \right\|_F^2,$$

and we can restrict our search to the case where the vectors a_i are orthogonal.

In $\widehat{\mathcal{V}}_r$, the objective function can be expanded as

$$(4.10) \quad \mathcal{O}_j(\mathbf{a}, \mathbf{b}) = \left\| S_j^\top R_j \right\|_F^2 + \sum_{i=1}^r \|a_i\|^2 \|C_j^\top b_i\|^2 + 2 \sum_{i=1}^r \text{tr}((S_j^\top R_j)^\top a_i b_i^\top C_j).$$

Imposing the gradient of (4.10) to vanish leads to decoupled equations for each pair (a_i, b_i) . In particular, for each $i = 1, \dots, r$, the resulting problem boils down to the rank-1 minimization of [24, Lemma 3.5]. For a fixed index i , a_i is given by (4.8), and b_i is eigenvector of the generalized eigenvalue problem

$$C_j (S_j^\top R_j)^\top (S_j^\top R_j) C_j^\top v = \lambda C_j C_j^\top v.$$

The corresponding eigenvalue $\lambda_i \in \mathbb{R}$ can be written as

$$\lambda_i = \frac{\left\| S_j^\top R_j C_j^\top b_i \right\|^2}{\|C_j^\top b_i\|^2}.$$

So far we know that if b_1, \dots, b_r are *any* r eigenvectors of (4.7) and a_1, \dots, a_r are obtained from b_i using (4.8), then $(\mathbf{a}, \mathbf{b}) \in \widehat{\mathcal{V}}_r$ is a stationary point of (4.6). We now want to identify the global minimum. Starting from (4.10) and inserting the expression for a_i , the second term on the right-hand side of (4.10) becomes

$$\sum_{i=1}^r \|a_i\|^2 \|C_j^\top b_i\|^2 = \sum_{i=1}^r \frac{\|S^\top R C_j^\top b_i\|^2}{\|C_j^\top b_i\|^2} = \sum_{i=1}^r \lambda_i,$$

while the third term of (4.10) is

$$\sum_{i=1}^r \text{tr}((S^\top R)^\top a_i b_i^\top C_j) = \sum_{i=1}^r -\frac{1}{\|C_j^\top b_i\|^2} \text{tr}(b_i^\top C_j (S^\top R)^\top (S^\top R) C_j^\top b_i) = -\sum_{i=1}^r \lambda_i.$$

Hence, the objective function (4.10) in $\widehat{\mathcal{V}}_r$ satisfies (4.9). Since \mathbf{b} is required to have column-rank r , the global minimum of (4.6) is obtained by choosing as b_1, \dots, b_r the eigenvectors of (4.7) corresponding to the r largest eigenvalues. \square

If the matrix C_j does not have full row-rank, that is, $r_C := \text{rank}(C_j) < m$, one can proceed as suggested in [24, Lemma 3.4] by performing a rank-revealing QR decomposition of C_j , i.e., $C_j = Q_j Z_j$ where $Q_j \in \mathbb{R}^{m \times r_C}$ has orthogonal columns and $Z_j \in \mathbb{R}^{r_C \times \overline{w}}$. Introducing $\mathbf{z} = Q_j^\top \mathbf{b}$, Theorem 4.1 gives the matrices $(\mathbf{a}, \mathbf{z}) \in \mathcal{V}_r(m_s, r_C)$ that minimize $\|S_j^\top R_j + \mathbf{a} \mathbf{z}^\top Z_j\|_F^2$, and \mathbf{b} can be recovered as $\mathbf{b} = Q_j \mathbf{z}$.

The steps to update the DEIM basis are summarized in Algorithm 4.1.

Algorithm 4.1 Adaptation of the DEIM projection

- 1: **procedure** $(U_{j+1}, P_{j+1}) = \text{ADAPTBASIS}(\{z^{\ell_{j+1}-w+1}, \dots, z^{\ell_{j+1}}\}, A, U_j, P_j, S_j, r)$
 - 2: Compute the matrix F_j at the interpolation and sampling points: $P_j^\top F_j, S_j^\top F_j$.
 - 3: Compute the DEIM coefficients $C_j = (P_j^\top U_j)^{-1} (P_j^\top F_j)$.
 - 4: Compute the DEIM residual at the sampling points: $S_j^\top R_j = S_j^\top U_j C_j - S_j^\top F_j$.
 - 5: Solve the generalized eigenvalue problem (4.7).
 - 6: Set b_1, \dots, b_r as the eigenvectors corresponding to the r largest eigenvalues.
 - 7: Compute a_1, \dots, a_r using (4.8).
 - 8: Set $\mathbf{a}_j = [a_1 \dots a_r]$ and $\mathbf{b}_j = [b_1 \dots b_r]$.
 - 9: $S_j^\top U_{j+1} \leftarrow S_j^\top U_j + \mathbf{a}_j \mathbf{b}_j^\top$ and $\check{S}_j^\top U_{j+1} \leftarrow \check{S}_j^\top U_j$.
 - 10: Compute the DEIM indices P_{j+1} associated with U_{j+1} .
 - 11: **end procedure**
-

For each value of the parameter, the complexity of updating the DEIM basis, as described in Algorithm 4.1, is $O(m_s k w (s_1 + s_2)) + O(m^2 m_s) + O(m_s m k w)$. In greater detail, under the assumption that the Jacobian $J_{\eta, h}$ has at most $s_1 + s_2$ non-zero entries per row, the computation of $P_j^\top F_j = P_j^\top J_{\eta, h}(A z^i) A$ for all w hyper-reduced solutions z^i in the temporal window has complexity $O((s_1 + s_2) m k w)$. Analogously, $S_j^\top F_j \in \mathbb{R}^{m_s \times \overline{w}}$ requires $O((s_1 + s_2) m_s k w)$ operations. Next, the matrix-matrix multiplications involved in the computation of $C_j \in \mathbb{R}^{m \times \overline{w}}$ (line 3) and of $S_j^\top R_j \in \mathbb{R}^{m_s \times \overline{w}}$ (line 4) require $O(m^3) + O(m^2 k w) + O(m_s m k w)$ operations. The solution of the generalized eigenvalue problem (4.7) has arithmetic complexity $O(m_s m k w) + O(m^2 m_s) + O(m^3)$. Finally, the total complexity of computing the matrix $\mathbf{a} \in \mathbb{R}^{m_s \times r}$ via (4.8) is $O(r m m_s + r m k w)$. The update of the DEIM interpolation points (line 11) can be performed, for example, by means of the greedy algorithm described in [9]

with complexity is $O(m^3) + O(m^2d)$, cf. [11]. This cost is linear in d because, at each iteration of [9, Algorithm 1], the norm of the d -dimensional DEIM residual needs to be computed. To reduce the computational burden of this step, one can update only the indices associated with the DEIM basis vectors that have undergone the largest rotations in the DEIM basis update from U_j to U_{j+1} , as suggested in [24, Section 4.1].

4.2. Sampling points update. A crucial aspect of Algorithm 4.1 is the definition of the sampling matrix S_j . To determine the best possible choice of sampling points, we analyze the reduction in the residual associated with the DEIM update and the distance of the DEIM space to the best approximation space at a given time. To this end, we first show that the DEIM update minimizes the projection error of the residual at the sampling points onto the space spanned by the rows of C_j .

LEMMA 4.2. *Let \bar{r} be the rank of the matrix $S_j^\top R_j C_j^\top \in \mathbb{R}^{m_s \times m}$ and assume that the DEIM coefficient matrix $C_j \in \mathbb{R}^{m \times \bar{w}}$ has full row-rank. Let U_{j+1} be the rank- \bar{r} update given by Algorithm 4.1. If $\mathbb{C}_j = C_j^\top (C_j C_j^\top)^{-1} C_j \in \mathbb{R}^{\bar{w} \times \bar{w}}$, then*

$$\|S_j^\top (U_{j+1} C_j - F_j)\|_F = \|S_j^\top R_j (I - \mathbb{C}_j)\|_F.$$

Proof. For all $i = 1, \dots, \bar{r}$, the eigenpairs (λ_i, b_i) satisfy the generalized eigenvalue problem (4.7), i.e. $C_j (S_j^\top R_j)^\top (S_j^\top R_j) C_j^\top \mathbf{b} = C_j C_j^\top \mathbf{b} \Lambda$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{\bar{r}})$. Since C_j has full row-rank, the matrix $C_j C_j^\top$ is symmetric and positive definite, thus admitting the Cholesky decomposition $C_j C_j^\top = L_j L_j^\top$, where $L_j \in \mathbb{R}^{m \times m}$ is lower triangular. We introduce the change of variables $\mathbf{c} = L_j^\top \mathbf{b}$, so that the eigenvalue problem becomes $L_j^{-1} C_j (S_j^\top R_j)^\top (S_j^\top R_j) C_j^\top L_j^{-\top} \mathbf{c} = \mathbf{c} \Lambda$. This means that the eigenvalues of the generalized problem (4.7) coincide with the squared singular values of the matrix $M_j := S_j^\top R_j C_j^\top L_j^{-\top} \in \mathbb{R}^{m_s \times m}$. Then, equation (4.9) can be written as

$$(4.11) \quad \|S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j\|_F^2 = \|S_j^\top R_j\|_F^2 - \sum_{\ell=1}^{\bar{r}} \sigma_\ell^2,$$

where $\sigma_1 \geq \dots \geq \sigma_{\bar{r}} > 0$ are the singular values of M_j . Since $\bar{r} = \text{rank}(S_j^\top R_j C_j^\top L_j^{-\top})$,

$$\|S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j\|_F^2 = \|S_j^\top R_j\|_F^2 - \|S_j^\top R_j C_j^\top L_j^{-\top}\|_F^2.$$

Moreover, by the cyclic property of the trace, it holds

$$\|S_j^\top R_j\|_F^2 - \|S_j^\top R_j C_j^\top L_j^{-\top}\|_F^2 = \text{tr}((I - C_j^\top (C_j C_j^\top)^{-1} C_j) (S_j^\top R_j)^\top (S_j^\top R_j)).$$

Since $\mathbb{C}_j = C_j^\top$ and $\mathbb{C}_j^2 = C_j$, we have $I - \mathbb{C}_j = (I - C_j)(I - C_j)^\top$, and

$$\|S_j^\top R_j + \mathbf{a} \mathbf{b}^\top C_j\|_F^2 = \text{tr}((I - C_j)(I - C_j)^\top (S_j^\top R_j)^\top (S_j^\top R_j)),$$

which concludes the proof. \square

The number of non-zero eigenvalues in the generalized eigenvalue problem (4.7) coincides with the number \bar{r} of non-zero singular values of $S_j^\top R_j C_j^\top L_j^{-\top}$. Furthermore, owing to (4.11), including eigenvectors of (4.7) corresponding to zero eigenvalues as columns of \mathbf{b} will have no effect on the value of the objective function (4.6). This suggests that the rank r of the update should be chosen so that $r \leq \bar{r}$.

In the following result we study the distance between the DEIM space resulting from the update described in Algorithm 4.1 and the best possible approximation space,

namely the one spanned by the columns of F_j , for $j = 0, \dots, N_a$. The resulting bound gives us an indication on how the sampling matrix can be chosen to minimize the error. Note that the result in the following theorem is analogous to [23, Proposition 2].

THEOREM 4.3. *Assume that $C_j \in \mathbb{R}^{m \times \bar{w}}$ has full row-rank. Let $U_{j+1} = U_j + \alpha_j \beta_j^\top$ be the rank- \bar{r} update of U_j obtained with [Algorithm 4.1](#). Let \mathcal{U}_{j+1} be the space spanned by the columns of U_{j+1} , and let the columns of F_j belong to the m -dimensional space $\bar{\mathcal{U}}_{j+1}$. Then, the distance between $\bar{\mathcal{U}}_{j+1}$ and \mathcal{U}_{j+1} is bounded as*

$$(4.12) \quad d(\bar{\mathcal{U}}_{j+1}, \mathcal{U}_{j+1}) := \|\bar{U}_{j+1} - U_{j+1} U_{j+1}^\top \bar{U}_{j+1}\|_F^2 \leq \frac{\rho_j^2}{\sigma_{\min}^2(F_j)}$$

where $\sigma_{\min}(F_j)$ is the smallest nonzero singular value of F_j and

$$\rho_j^2 := \|R_j(I - \mathbb{C}_j)\|_F^2 + \|\check{S}_j^\top R_j \mathbb{C}_j\|_F^2 = \|R_j\|_F^2 - \|S_j^\top R_j \mathbb{C}_j\|_F^2.$$

Proof. First note that the update $\alpha_j \beta_j^\top$ only changes the rows of U_j corresponding to the sampling points S_j as shown in (4.5). [Lemma 4.2](#) yields

$$(4.13) \quad \begin{aligned} \|U_{j+1} C_j - F_j\|_F^2 &= \|S_j^\top R_j(I - \mathbb{C}_j)\|_F^2 + \|\check{S}_j^\top R_j\|_F^2 \\ &= \|R_j(I - \mathbb{C}_j)\|_F^2 - \|\check{S}_j^\top R_j(I - \mathbb{C}_j)\|_F^2 + \|\check{S}_j^\top R_j\|_F^2. \end{aligned}$$

The linearity and cyclic property of the trace, together with the fact that \mathbb{C}_j is a projection, give $\|\check{S}_j^\top R_j\|_F^2 - \|\check{S}_j^\top R_j(I - \mathbb{C}_j)\|_F^2 = \|\check{S}_j^\top R_j \mathbb{C}_j\|_F^2$. Hence, equation (4.13) yields $\|U_{j+1} C_j - F_j\|_F^2 = \rho_j^2$.

Let the columns of \bar{U}_{j+1} form an orthonormal basis of $\bar{\mathcal{U}}_{j+1}$. Since the columns of F_j belong to $\bar{\mathcal{U}}_{j+1}$, there is a matrix $\tilde{F}_j \in \mathbb{R}^{m \times \bar{w}}$ such that $F_j = \bar{U}_{j+1} \tilde{F}_j$. Hence,

$$(4.14) \quad \begin{aligned} \|F_j - U_{j+1} U_{j+1}^\top F_j\|_F^2 &= \|(\bar{U}_{j+1} - U_{j+1} U_{j+1}^\top \bar{U}_{j+1}) \tilde{F}_j\|_F^2 \\ &\geq \|\bar{U}_{j+1} - U_{j+1} U_{j+1}^\top \bar{U}_{j+1}\|_F^2 \sigma_{\min}^2(\tilde{F}_j), \end{aligned}$$

and $\sigma_{\min}(\tilde{F}_j) = \sigma_{\min}(F_j)$ since \bar{U}_{j+1} is orthonormal. The conclusion follows by combining (4.14) with the bound $\|F_j - U_{j+1} U_{j+1}^\top F_j\|_F^2 \leq \|F_j - U_{j+1} C_j\|_F^2$. \square

If $\rho_j = 0$ in (4.12), then the spaces \mathcal{U}_{j+1} and $\bar{\mathcal{U}}_{j+1}$ coincide, and the adapted DEIM basis U_{j+1} can exactly represent all snapshots F_j of the reduced Jacobian in the window. This observation suggests that the ideal set of sample indices S_j is the one that minimizes ρ_j . In particular, since R_j and \mathbb{C}_j are fixed, we choose the S_j that minimizes $\|\check{S}_j^\top R_j \mathbb{C}_j\|_F$ or, equivalently, maximizes $\|S_j R_j \mathbb{C}_j\|_F$, as described in [Algorithm 4.2](#). We remark the adaptive method proposed in [23] is based on a different update of the DEIM basis, performed via SVD of the residual matrix R_j (see [23, Algorithm 3]), and the sampling matrix S_j is chosen to minimize $\|\check{S}_j^\top R_j\|_F$.

Algorithm 4.2 Adaptation of the sampling indices

- 1: **procedure** $S_{j+1} = \text{ADAPTSAMPLEINDICES}(\{z^{\ell_{j+1}-w+1}, \dots, z^{\ell_{j+1}}\}, A, U_j, P_j, m_s)$
 - 2: Build the snapshot matrix F_j of the reduced Jacobian in the window.
 - 3: Compute $R_j \mathbb{C}_j = U_j C_j - F_j C_j^\top (C_j C_j^\top)^{-1} C_j$.
 - 4: Take as new sampling indices the indices of the m_s rows of $R_j \mathbb{C}_j$ with largest norm.
 - 5: **end procedure**
-

To analyze the computational cost of [Algorithm 4.2](#), we first observe that some of the quantities involved are already available from the basis update ([Algorithm 4.1](#)). The operations required are the computation of the $d - m$ entries of F_j not available from $P_j^\top F_j$, at a cost of order $O((d - m)kw(s_1 + s_2))$. To construct \mathbb{C}_j one can use the SVD of $C_j = U_j \Sigma_j \mathcal{V}_j^\top$ so that $\mathbb{C}_j = \tilde{\mathcal{V}}_j \tilde{\mathcal{V}}_j^\top$, where $\tilde{\mathcal{V}}_j \in \mathbb{R}^{\bar{w} \times m}$ is obtained by selecting the first m columns of \mathcal{V}_j . In this case, $R_j \mathbb{C}_j = U_j C_j - F_j \tilde{\mathcal{V}}_j \tilde{\mathcal{V}}_j^\top$ can be computed with $O(dmkw)$ operations. The arithmetic complexity of [Algorithm 4.2](#) is then linear in d . Since d typically scales with the full dimension n (see discussion in [Remark 3.1](#)), the update of the sampling indices turns out to be too expensive to be performed at each adaptation step. In particular, solving the adaptive hyper-reduced system may become computationally demanding for large values of w or m , as observed in the numerical experiments, [Table 4](#) and [Figure 7](#) in particular. For this reason, it is preferable to perform the update of the sampling indices every γ steps, where $\gamma = \nu\delta$ and $\nu > 1$. Moreover, instead of being fixed *a priori*, the number of sampling indices m_s may be determined at each adaptation until $\|\tilde{S}_j R_j \mathbb{C}_j\|_F$ is below a threshold tolerance. We refer to [Subsection 5.2](#) for a numerical study regarding these aspects.

The gradient-preserving adaptive DEIM scheme is summarized in [Algorithm 4.3](#).

Algorithm 4.3 Gradient-preserving adaptive DEIM

```

1: procedure GP-ADEIM( $A, \delta_0, \delta, w, \gamma, m_s, r$ )
2:   Compute  $(U_0, P_0)$  from snapshots  $J_{\eta, h}(AA^\top y^\ell(\eta))$  for  $\ell = 0, \dots, \delta_0$ .
3:    $j \leftarrow 0$ 
4:   for  $\tau = 1, \dots, N_t - 1$  do
5:     Compute  $z^\tau$  by solving (4.1) with  $U_j$  and  $P_j$ .
6:     if  $\text{mod}(\tau - \delta_0, \delta) = 0$  then
7:       Collect reduced trajectories in the window:  $\mathcal{Z} = \{z^{\delta_0 + j\delta - w + 1}, \dots, z^{\delta_0 + j\delta}\}$ .
8:       if  $\text{mod}(\tau - \delta_0, \gamma) = 0$  then
9:          $S_j \leftarrow \text{ADAPTSAMPLEINDICES}(\mathcal{Z}, A, U_j, P_j, m_s)$  with Algorithm 4.2.
10:      else
11:         $S_j \leftarrow S_{j-1}$ 
12:      end if
13:       $(U_{j+1}, P_{j+1}) \leftarrow \text{ADAPTBASIS}(\mathcal{Z}, A, U_j, P_j, S_j, r)$  with Algorithm 4.1.
14:       $j \leftarrow (\tau - \delta_0) / \delta + 1$ 
15:    end if
16:  end for
17: end procedure

```

4.3. Conservation of the Hamiltonian. Similarly to the study conducted in [Subsection 3.3](#) for the non-adaptive hyper-reduction, we assess the error in the conservation of the Hamiltonian due to its approximation via *local* hyper-reduction. In the following result we show that the error between the Hamiltonian evaluated at the full model solution and at the hyper-reduced solution at a given time is bounded by the sum of the local hyper-reduction error of h at all previous updates.

PROPOSITION 4.4. *Let $\eta \in \mathcal{P}$ be fixed. Let $y^{\ell_j}(\eta)$ be an approximation of the solution $y(t^{\ell_j}, \eta)$ of the full order system (2.1) at time t^{ℓ_j} , with $j = 1, \dots, N_a + 1$, obtained with a user-defined numerical time integrator. Similarly, let $z^{\ell_j}(\eta)$ be an approximation of the solution $z(t^{\ell_j}, \eta)$ of the hyper-reduced system (3.5) at time t^{ℓ_j} . Assume that $y^0(\eta) \in \text{Col}(A)$ and $h(AA^\top y^0(\eta), \eta) \in \text{Col}(U_0)$. Furthermore, assume that the basis is updated so that $h(Az^{\ell_j}(\eta), \eta) \in \text{Col}(U_j)$ for all $j = 1, \dots, N_a$. Then,*

the error $\Delta\mathcal{H}_j(\eta) := |\mathcal{H}(y^{\ell_j}(\eta), \eta) - \mathcal{H}(Az^{\ell_j}(\eta), \eta)|$ satisfies

$$(4.15) \quad \Delta\mathcal{H}_j(\eta) \leq \sum_{i=0}^{j-1} (|v^\top (I - \mathbb{P}_i)h(Az^{\ell_{i+1}}(\eta), \eta)| + \varepsilon_{\mathcal{H}_{hr}^i}^{[t^{\ell_i}, t^{\ell_{i+1}}]}) + \varepsilon_{\mathcal{H}}^{[t^0, t^{\ell_j}]},$$

where $\varepsilon_{\mathcal{H}_{hr}^j}^{[t^{\ell_j}, t^{\ell_{j+1}}]} := |\mathcal{H}_{hr}^j(z^{\ell_{j+1}}) - \mathcal{H}_{hr}^j(z^{\ell_j})|$ and $\varepsilon_{\mathcal{H}}^{[t^0, t^{\ell_j}]} := |\mathcal{H}(y^{\ell_j}) - \mathcal{H}(y^0)|$ are the errors in the Hamiltonian conservation in the specified interval and associated with the chosen temporal integrator.

Proof. In this proof we omit the dependence of y^{ℓ_j} , z^{ℓ_j} , and of \mathcal{H} on η . The Hamiltonian error at the generic time instant t^{ℓ_j} can be bounded as

$$\Delta\mathcal{H}_j(\eta) \leq |\mathcal{H}(y^{\ell_j}) - \mathcal{H}(y^0)| + |\mathcal{H}(y^0) - \mathcal{H}(Az^0)| + |\mathcal{H}(Az^0) - \mathcal{H}(Az^{\ell_j})|.$$

The first term on the right-hand side is $\varepsilon_{\mathcal{H}}^{[t^0, t^{\ell_j}]}$ and only depends on the time integration scheme. Moreover, the second term is zero since $y^0 \in \text{Col}(A)$ by assumption. As shown in [Subsection 3.3](#), this assumption can be easily enforced by a suitable shift of the initial condition. The third term can be bounded as

$$\begin{aligned} |\mathcal{H}(Az^0) - \mathcal{H}(Az^{\ell_j})| &\leq |\mathcal{H}(Az^0) - \mathcal{H}_{hr}^0(z^0)| + \sum_{i=0}^{j-1} |\mathcal{H}_{hr}^i(z^{\ell_{i+1}}) - \mathcal{H}_{hr}^i(z^{\ell_i})| \\ &\quad + \sum_{i=0}^{j-2} |\mathcal{H}_{hr}^{i+1}(z^{\ell_{i+1}}) - \mathcal{H}_{hr}^i(z^{\ell_{i+1}})| + |\mathcal{H}_{hr}^{j-1}(z^{\ell_j}) - \mathcal{H}(Az^{\ell_j})|. \end{aligned}$$

The first term on the right-hand side vanishes if $h(Az^0) \in \text{Col}(U_0)$. Furthermore, the quantity $|\mathcal{H}_{hr}^i(z^{\ell_i}) - \mathcal{H}_{hr}^i(z^{\ell_{i+1}})| = \varepsilon_{\mathcal{H}_{hr}^i}^{[t^{\ell_i}, t^{\ell_{i+1}}]}$ only depends on the temporal integrator. The third term can be further split as follows, for all $i = 0, \dots, j-2$,

$$|\mathcal{H}_{hr}^i(z^{\ell_{i+1}}) - \mathcal{H}_{hr}^{i+1}(z^{\ell_{i+1}})| \leq |\mathcal{H}_{hr}^i(z^{\ell_{i+1}}) - \mathcal{H}(Az^{\ell_{i+1}})| + |\mathcal{H}(Az^{\ell_{i+1}}) - \mathcal{H}_{hr}^{i+1}(z^{\ell_{i+1}})|,$$

where the second term vanishes if $h(Az^{\ell_{i+1}}) \in \text{Col}(U_{i+1})$.

The conclusion follows from the definition of the Hamiltonians involved. \square

5. Numerical experiments. To numerically assess the performances of the proposed methods, we consider two test problems: the 2D shallow water equations and the 1D nonlinear Schrödinger equation. As numerical time integration schemes we compare the implicit midpoint rule (IMR), which is a symplectic time integrator [15], and the Average Vector Field (AVF), which is not symplectic but exactly preserves the Hamiltonian [26, 7]. The tolerance of the Newton method used within the implicit timestepping is set to 10^{-10} . In all the numerical tests, we evaluate the accuracy of the reduced and hyper-reduced solutions by means of the following relative errors:

$$(5.1) \quad \mathcal{E}_{L^2} = \sqrt{\frac{\sum_{i=0}^{N_t} \|y^i - Az^i\|^2}{\sum_{i=0}^{N_t} \|y^i\|^2}}, \quad \mathcal{E}_{\text{fin}} = \frac{\|y^{N_t} - Az^{N_t}\|}{\|y^{N_t}\|}.$$

We shall distinguish between $\mathcal{E}_{L^2}^R$, $\mathcal{E}_{\text{fin}}^R$ and $\mathcal{E}_{L^2}^{\text{HR}}$, $\mathcal{E}_{\text{fin}}^{\text{HR}}$ according to whether z^i is the reduced or the hyper-reduced solution at time t^i , respectively.

5.1. Two-dimensional shallow water equations. The two-dimensional shallow water equations (2D-SWE) [28] in the rectangular domain $\Omega = [-L_{x_1}, L_{x_1}] \times [-L_{x_2}, L_{x_2}] \subset \mathbb{R}^2$ reads

$$(5.2) \quad \begin{cases} \partial_t \chi + \gamma \nabla \cdot (\chi \nabla \Phi), & \text{in } \Omega \times (0, T], \\ \partial_t \Phi + \frac{\gamma}{2} |\nabla \Phi|^2 + \gamma \chi = 0, & \text{in } \Omega \times (0, T], \end{cases}$$

where ∂_t is short-hand notation for the temporal derivative, $\nabla \cdot$ and ∇ denote the divergence and gradient operators in $\mathbf{x} = (x_1, x_2) \in \Omega$, respectively; $\chi = \chi(\mathbf{x}, t; \eta)$ is the height of the free surface and $\Phi = \Phi(\mathbf{x}, t; \eta)$ is the scalar velocity potential, that is, $\nabla \Phi$ is the horizontal velocity normalized by the characteristic speed; $\gamma \in \mathbb{R}$ is a constant-valued parameter. We complement the problem with periodic boundary conditions and initial conditions. Problem (5.2) admits a Hamiltonian formulation with Hamiltonian

$$\widehat{\mathcal{H}}(\chi, \phi) = \frac{\gamma}{2} \int_{\Omega} (\chi |\nabla \phi|^2 + \chi^2) dx.$$

We discretize the spatial differential operators in (5.2) using centered second-order finite differences. To this end, we introduce a uniform Cartesian mesh on the domain Ω with $((x_1)_i, (x_2)_j) = (-L_{x_1} + i\Delta x_1, -L_{x_2} + j\Delta x_2)$, for $i = 0, \dots, n_{x_1}$, $j = 0, \dots, n_{x_2}$ and $\Delta x_1 = 2L_{x_1}/n_{x_1}$, $\Delta x_2 = 2L_{x_2}/n_{x_2}$. Let $n := n_{x_1}n_{x_2}$ and $D_{x_1}, D_{x_2} \in \mathbb{R}^{n \times n}$ be the matrices corresponding to the finite-difference discretization of the spatial derivatives. Defining the vectors $\boldsymbol{\chi}(t, \eta) = (\chi_1, \dots, \chi_n)^\top$, and $\boldsymbol{\Phi}(t, \eta) = (\Phi_1, \dots, \Phi_n)^\top$, approximating the nodal values of χ and Φ respectively, we obtain the discretized system

$$\begin{cases} \dot{\boldsymbol{\chi}} + \gamma [D_{x_1}(\boldsymbol{\chi} \odot D_{x_1} \boldsymbol{\Phi}) + D_{x_2}(\boldsymbol{\chi} \odot D_{x_2} \boldsymbol{\Phi})] = \mathbf{0}, \\ \dot{\boldsymbol{\Phi}} + \gamma \boldsymbol{\chi} + \frac{\gamma}{2} [(D_{x_1} \boldsymbol{\Phi})^2 + (D_{x_2} \boldsymbol{\Phi})^2] = \mathbf{0}. \end{cases}$$

Here, the symbol \odot denotes the pointwise product of vectors. This is a Hamiltonian system of the form (2.1) with $y(t, \eta) = (\boldsymbol{\chi}^\top(t, \eta), \boldsymbol{\Phi}^\top(t, \eta))^\top \in \mathbb{R}^{2n}$ and

$$\mathcal{N}(y, \eta) = \frac{\gamma}{2} \sum_{i=1}^n \chi_i [(D_{x_1} \boldsymbol{\Phi})_i^2 + (D_{x_2} \boldsymbol{\Phi})_i^2] = v^\top h(y, \eta),$$

where $v \in \mathbb{R}^n$ is the vector having all entries equal to one and $h(y, \eta) \in \mathbb{R}^n$ is defined as $(h(y, \eta))_i := \frac{\gamma}{2} \chi_i [(D_{x_1} \boldsymbol{\Phi})_i^2 + (D_{x_2} \boldsymbol{\Phi})_i^2]$ for all $i = 1, \dots, n$. Note that, in this case, the decomposition of \mathcal{N} has $d = n$.

In the following experiments we set $L_{x_1} = L_{x_2} = 2$ and take $n_{x_1} = n_{x_2} = 50$ grid points per direction, corresponding to $\Delta x_1 = \Delta x_2 = 0.08$ and $n = 2500$. The final time is $T = 10$ and the time step is $\Delta t = 0.005$. The initial condition is $\Phi(\mathbf{x}, 0) = 0$, while the vertical displacement χ is a Gaussian profile of the form $\chi(\mathbf{x}, 0; \eta) = 1 + \frac{1}{2} e^{-\beta |\mathbf{x}|^2}$. The problem parameter is $\eta = (\beta, \gamma)$ and it ranges in the set $\mathcal{P} = [1.1, 1.7] \times [0.7, 1.3] \subset \mathbb{R}^2$. The shift described in Subsection 3.3 is performed on the initial condition.

For this test case, we compare the performances of the full model, the reduced model and the hyper-reduced model obtained with the non-adaptive algorithm. To construct the reduced model, the full system is solved, in the offline phase, for 5 equispaced values of β and γ , for a total of 25 training parameters in the training

set \mathcal{P}_h . Next, the reduced basis is built via complex SVD using all N_t temporal snapshots for every training parameter. We choose two different dimensions of the reduced space, namely $2k = 80$ and $2k = 160$. For the hyper-reduced problem, we build the DEIM basis via POD on snapshots of the reduced Jacobian at every 20 time steps for each training parameter.

We test the reduced model (2.3) and the hyper-reduced model (3.5) for the randomly selected parameter $\eta = (1.6435, 0.7762) \notin \mathcal{P}_h$. In Figure 1 we report the relative errors (5.1) of the hyper-reduced model vs. the size m of the DEIM basis, obtained with the AVF and IMR time integrators and for the two sizes $2k \in \{80, 160\}$ of the reduced space. As expected, the error of the hyper-reduced model converges to the error of the reduced model as m increases. Moreover, as the size $2k$ of the reduced model increases, a larger number of DEIM bases are required to approach the limit value. The errors obtained with the two numerical time integration schemes are very similar.

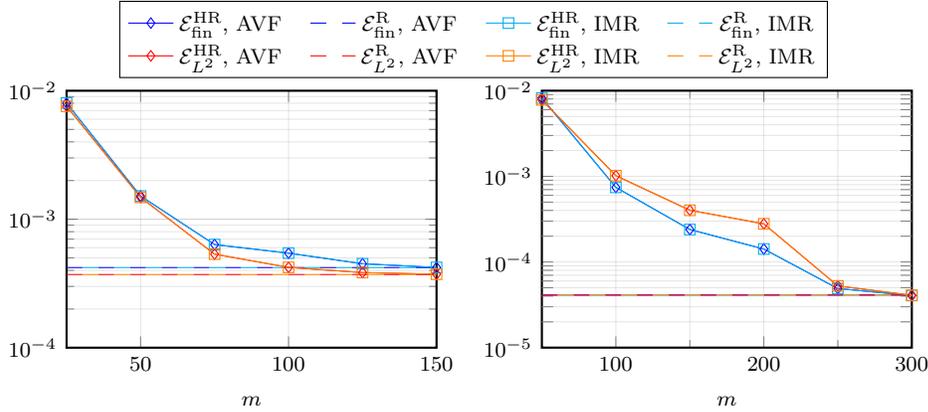


FIG. 1. 2D-SWE. Errors (5.1) of the reduced and hyper-reduced models vs. size m of the DEIM space. The size of the reduced model is $2k = 80$ (left) and $2k = 160$ (right). Comparison of AVF and IMR for time integration.

The computational times required by the reduced and hyper-reduced models solved using the AVF and the IMR time integrator are reported in Table 1. For $2k = 80$, solving the hyper-reduced system with, e.g., $m = 150$ is about 37 times faster than solving the reduced system for AVF, and about 34 times faster for IMR. For $2k = 160$, solving the hyper-reduced system with $m = 300$ is about 18 times faster than solving the reduced system for AVF, and about 15 times faster for IMR. Note that, in this test case, AVF is more computationally expensive than IMR, because an integral needs to be numerically evaluated at each time step.

We then consider the conservation of the Hamiltonian for a fixed parameter $\eta = (1.6435, 0.7762)$, whose dependence we omit in the following. In particular, we monitor the error $|\mathcal{H}(y^0) - \mathcal{H}(Az^j)|$ for all time indices $j = 0, \dots, N_t$. This quantity is bounded, in turn, by the error $\Delta\mathcal{H}_j$ in approximating the reduced Hamiltonian with the hyper-reduced one: as derived in Subsection 3.3, it holds

$$|\mathcal{H}(y^0) - \mathcal{H}(Az^j)| \leq \varepsilon_{\mathcal{H}}^{[t^0, t^j]} + \Delta\mathcal{H}_j \quad \text{for all } j = 0, \dots, N_t.$$

Figure 2 shows the three terms involved in the bound at every time instant t^j when the reduced space has dimension $2k = 160$. The error $\varepsilon_{\mathcal{H}}^{[t^0, t^j]}$ in the conservation of

TABLE 1

2D-SWE. Computational times of the online phases of the reduced and hyper-reduced systems, $2k = 80$ and $2k = 160$. The computational time required to solve the full order system is 414.10 s for AVF, 297.29 s for IMR. The reported times are computed as averages over 3 runs.

	Reduced model	Hyper-reduced model					
AVF	$2k = 80$	$m = 25$	$m = 50$	$m = 75$	$m = 100$	$m = 125$	$m = 150$
	276.02 s	2.78 s	3.86 s	5.20 s	5.81 s	7.10 s	7.44 s
	$2k = 160$	$m = 50$	$m = 100$	$m = 150$	$m = 200$	$m = 250$	$m = 300$
	307.02 s	7.44 s	9.85 s	12.16 s	13.55 s	16.08 s	17.12 s
IMR	$2k = 80$	$m = 25$	$m = 50$	$m = 75$	$m = 100$	$m = 125$	$m = 150$
	128.22 s	1.08 s	1.60 s	2.54 s	2.96 s	3.40 s	3.71 s
	$2k = 160$	$m = 50$	$m = 100$	$m = 150$	$m = 200$	$m = 250$	$m = 300$
	146.67 s	5.25 s	6.26 s	7.19 s	8.13 s	9.36 s	10.04 s

the full order Hamiltonian (black line in Figure 2) only depends on the timestepping and reaches the tolerance of the Newton solver for AVF, while it depends on Δt for IMR. The error in the conservation of the Hamiltonian associated with the hyper-reduced model decreases as m increases, that is, as the quality of the approximation of the Hamiltonian improves. For m sufficiently large, in this test case $m \geq 350$, the error due to the time integration is dominating over the error introduced by the Hamiltonian hyper-reduction. The reason is that the hyper-reduced model with $m \geq 350$ is as accurate as the reduced model and, hence, the Hamiltonian is exactly preserved up to the error of the temporal integrator. As a further confirmation of the relationship between the conservation and the hyper-reduction of the Hamiltonian stated in Proposition 3.4, we also plot the DEIM approximation error $|\mathcal{H}_{hr}(z^j) - \mathcal{H}(Az^j)| = |v^\top (\mathbb{P}h(Az^j) - h(Az^j))|$ for all $j = 0, \dots, N_t$.

$\cdots \square \cdots \Delta \mathcal{H}_j, m = 200$	$\cdots \diamond \cdots \mathcal{H}(y^0) - \mathcal{H}(Az^j) , m = 200$	$\cdots \triangle \cdots \mathcal{H}_{hr}(z^j) - \mathcal{H}(Az^j) , m = 200$
$\cdots \square \cdots \Delta \mathcal{H}_j, m = 250$	$\cdots \diamond \cdots \mathcal{H}(y^0) - \mathcal{H}(Az^j) , m = 250$	$\cdots \triangle \cdots \mathcal{H}_{hr}(z^j) - \mathcal{H}(Az^j) , m = 250$
$\cdots \square \cdots \Delta \mathcal{H}_j, m = 300$	$\cdots \diamond \cdots \mathcal{H}(y^0) - \mathcal{H}(Az^j) , m = 300$	$\cdots \triangle \cdots \mathcal{H}_{hr}(z^j) - \mathcal{H}(Az^j) , m = 300$
$\cdots \square \cdots \Delta \mathcal{H}_j, m = 350$	$\cdots \diamond \cdots \mathcal{H}(y^0) - \mathcal{H}(Az^j) , m = 350$	$\cdots \triangle \cdots \mathcal{H}_{hr}(z^j) - \mathcal{H}(Az^j) , m = 350$

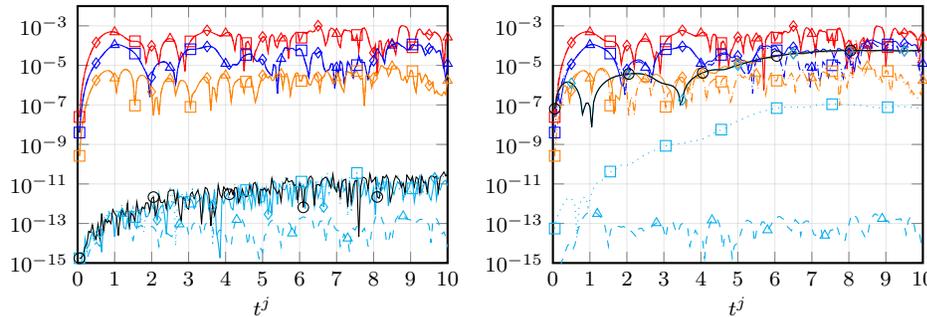


FIG. 2. 2D-SWE. Conservation of the Hamiltonian over time for different choices of m . The solid black lines with circles correspond to the error $\varepsilon_{\mathcal{H}}^{[t^0, t^j]}$ in the conservation of the full Hamiltonian due to the timestepping. AVF (left) and IMR (right) for time integration. The reduced model has dimension $2k = 160$ and the test parameter is $\eta = (1.6435, 0.7762)$.

5.2. Nonlinear Schrödinger equation. As a second test case, we consider the nonlinear Schrödinger equation (NLS)

$$(5.3) \quad i \partial_t u + \partial_{xx} u + \epsilon |u|^2 u = 0, \quad x \in [-L, L], t \in (0, T],$$

with initial condition $u(x, 0; \eta) = \sqrt{2}(\cosh x)^{-1} \exp(i\frac{x}{2})$ and periodic boundary conditions. Here ∂_{xx} denotes the second order derivative with respect to the spatial variable, and $\epsilon = \eta \in \mathbb{R}$ is a parameter ranging in the set $\mathcal{P} = [0.9, 1.1]$. Writing $u(x, t; \eta) = q(x, t; \eta) + ip(x, t; \eta)$, for all $(x, t) \in [-L, L] \times [0, T]$, problem (5.3) admits a Hamiltonian formulation with Hamiltonian given by

$$\widehat{\mathcal{H}}(q, p) = \frac{1}{2} \int_{-L}^L \left((\partial_x p)^2 + (\partial_x q)^2 - \frac{\epsilon}{2} (q^2 + p^2)^2 \right) dx.$$

Consider the uniform computational grid $x_i = -L + i\Delta x$, for $i = 0, \dots, n$ and $\Delta x = 2L/n$. Let $D_{xx} \in \mathbb{R}^{n \times n}$ be the matrix corresponding to the central finite-difference discretization of the second order spatial derivative. Introducing the vectors $\mathbf{q}(t, \eta) = (q_1, \dots, q_n)^\top$ and $\mathbf{p}(t, \eta) = (p_1, \dots, p_n)^\top$, where q_i and p_i are approximations to $q(x_i)$ and $p(x_i)$, we derive the system

$$\begin{cases} \dot{\mathbf{q}} = -D_{xx}\mathbf{p} - \epsilon(\mathbf{q}^2 + \mathbf{p}^2) \odot \mathbf{p}, \\ \dot{\mathbf{p}} = D_{xx}\mathbf{q} + \epsilon(\mathbf{q}^2 + \mathbf{p}^2) \odot \mathbf{q}, \end{cases}$$

which is of the form (2.1) with $y(t, \eta) = (\mathbf{q}^\top(t, \eta), \mathbf{p}^\top(t, \eta))^\top \in \mathbb{R}^{2n}$ and

$$\mathcal{N}(y) = -\frac{\epsilon}{4} \sum_{i=1}^n (q_i^2 + p_i^2)^2 = v^\top h(y).$$

Here $v \in \mathbb{R}^n$ is the vector with entries all equal to one and $h(y) \in \mathbb{R}^n$ is defined as $(h(y))_i := -\frac{\epsilon}{4}(q_i^2 + p_i^2)^2$ for all $i = 1, \dots, n$. In the following tests we set $L = \pi/l$ with $l = 0.11$ as in [1] and $n = 2048$. AVF time integrator is applied in the temporal interval $[0, 30]$ with time step $\Delta t = 0.01$.

The reduced basis is built using the complex SVD on snapshots of the full model solution at each time step and for $|\mathcal{P}_h| = 11$ equispaced training parameters in the interval $\mathcal{P} = [0.9, 1.1]$. Each simulation of the full model takes about 220 seconds.

Before studying the performances of the proposed hyper-reduction technique, we compare it with the SDEIM proposed in [1] and [25]. SDEIM consists in applying a DEIM approximation to the nonlinear Hamiltonian gradient and to chose as DEIM basis the symplectic basis used to approximate the state, constructed by adding suitable snapshots to improve accuracy. The resulting method does not exactly preserve the gradient structure of the Hamiltonian vector field, as pointed out in [25, page A17] and in [17, page 1717], but it ensures *asymptotic* boundedness of the energy of the system [25, Theorem 5.1]. However, this property is not enough to guarantee stability of the resulting approximate solution. To illustrate these facts, we compare our hyper-reduction technique (non-adaptive version) with SDEIM tested on the Schrödinger equation with $\eta = 1.0932$. Since in the SDEIM algorithm the DEIM basis is chosen to be the same as the reduced basis, we set $m = 2k$ in the gradient-preserving hyper-reduction method so that the reduced and DEIM spaces have the same dimension for the two methods. In particular we consider $k = m/2 \in \{10, 20, 30, 40, 50, 60\}$. To build the reduced and DEIM spaces, we consider all the snapshots of the full solution from the training phase, and one snapshot every $d_{snap} = 20$ time steps for the

nonlinear term. Figure 3 shows a pseudocolor plot of the numerical solutions on a space-time domain: The solution produced by SDEIM is unstable for $2k = 80$ (first row) and inaccurate for $2k = 120$ (second row). We also point out that, similarly, when classical DEIM is used to approximate the nonlinear term unstable solutions and unbounded errors are produced.

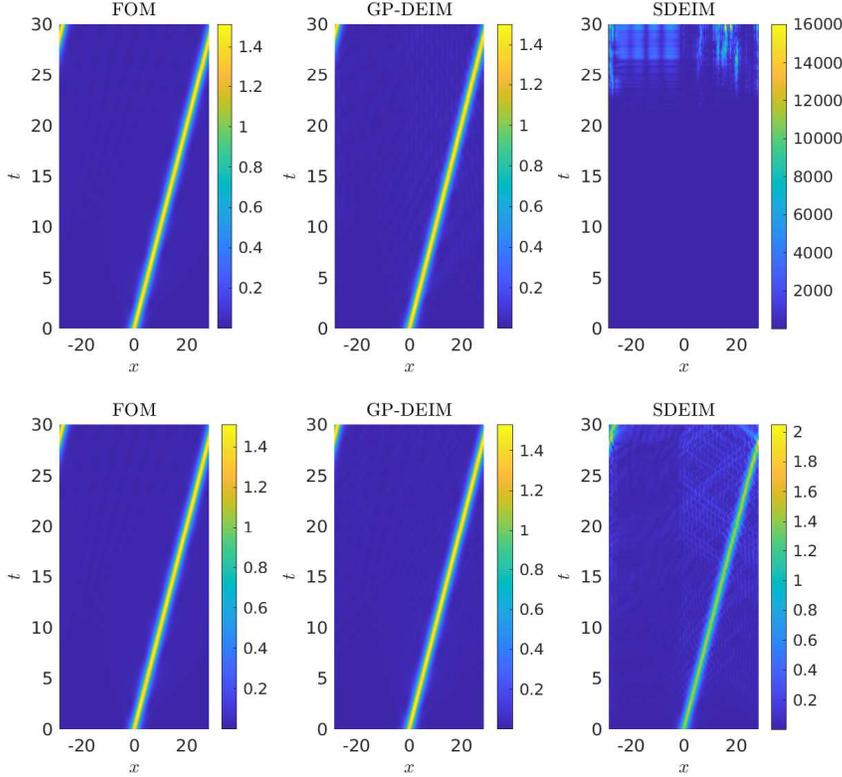


FIG. 3. Numerical solution of the Schrödinger equation in space-time with $\eta = 1.0932$, for $2k = m = 80$ (first row) and $2k = m = 120$ (second row). The first column refers to the full order solution, the second column to the proposed gradient-preserving hyper-reduction, and the third column to the SDEIM of [25, 1].

In this problem, the dimension of the DEIM basis required by the hyper-reduced model to achieve the error of the reduced model can be rather large. This behavior can be ascribed to the fact that the Schrödinger equation and its Hamiltonian do not exhibit significant global reducibility properties. Figure 4 shows, on the left, the singular values of the full order solution obtained for $N_t = 3000$ time steps and, on the right, the singular values of the snapshot matrix $M_J \in \mathbb{R}^{n \times 2kN_s}$ of the Jacobian for $N_s = 150$ time instants (one every 20 time steps). One value of the parameter is considered: $\eta = 1.0932$. Each line corresponds to a different dimension $2k$ of the reduced model. We can observe that the singular values of the snapshots corresponding to the full order solution decay relatively slowly (Figure 4, left), which suggests that the dimension $2k$ of the reduced space needs to be sufficiently large to accurately represent the full dynamics. On the other hand, the reducibility of the Jacobian deteriorates as $2k$ increases (Figure 4, right). This last remark also justifies our choice of performing hyper-reduction on the reduced Jacobian rather than on the full order

one, as already pointed out in [Remark 3.2](#).

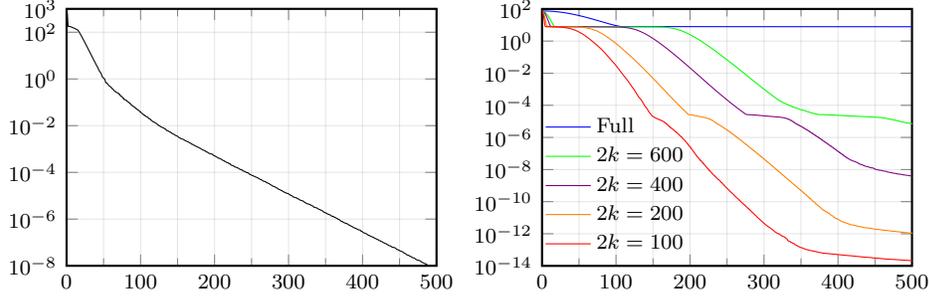


FIG. 4. NLS. Left: singular values of the snapshot matrix of the full order solution. Right: singular values of the snapshot matrix $M_J \in \mathbb{R}^{n \times 2nN_s}$ of the Jacobian for different choices of k . Only the 500 largest singular values are shown.

To address the lack of global reducibility of the Jacobian, we perform the adaptation of DEIM basis and indices described in [Section 4](#). In this section we fix the size of the reduced basis to $2k = 400$ and we consider $\eta = 1.0932 \notin \mathcal{P}_h$ as test parameter.

As a first test, we consider, at each update $j = 1, \dots, N_a$, the selection of the m_s sampling points collected in the matrix S_j and we compare three possible selection strategies: (i) random; (ii) based on the minimization of the residual norm $\|R_j\|_F$ as proposed in [\[23, Algorithm 1\]](#); and (iii) based on the minimization of the projected residual norm $\|R_j C_j\|_F$ motivated by [Theorem 4.3](#), see also [Algorithm 4.2](#). We set the dimension of the DEIM basis to $m = 100$, while the adaptation hyper-parameters are $\delta = \delta_0 = \gamma = 5$, $w = 1$ and $r = \bar{r}$, namely a full-rank update is performed at each step. We compare the three selection strategies for two possible sizes of the sampling matrix, namely $m_s \in \{150, 300\}$. [Figure 5](#) shows the error $\|U_{j+1} C_j - F_j\|_F$, at each update j , for the three selection criteria. It can be observed that the random algorithm gives the worst results, while the other two strategies do not yield significantly different results, at least in this simulation.

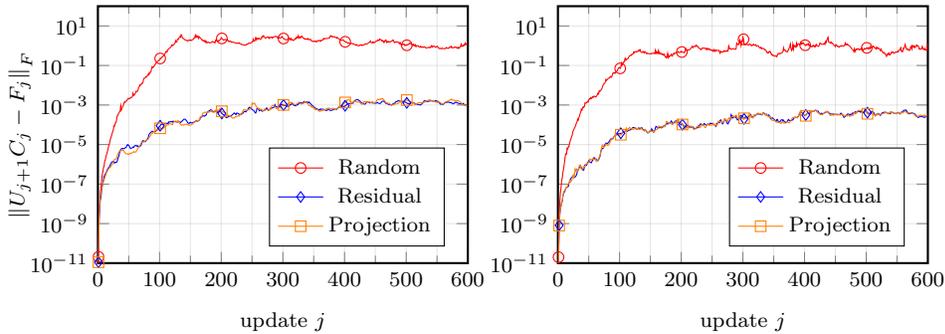


FIG. 5. NLS. Evolution of $\|U_{j+1} C_j - F_j\|_F$ for three different criteria of selection of the sampling points. The number of sampling points is $m_s = 150$ (left plot) and $m_s = 300$ (right plot). The dimension of the reduced basis is $2k = 400$ and the dimension of the DEIM basis is $m = 100$.

Next, we assess the adaptive hyper-reduction algorithm for different choices of the adaptation hyper-parameters – r , m_s , w , δ , and γ – in terms of: (i) accuracy of the solution, measured by the error $\mathcal{E}_{L^2}^{\text{HR}}$ from [\(5.1\)](#); (ii) computational time of

both the offline phase, t_{off} , and the online phase, t_{on} , computed as the average time, in seconds, over 5 independent runs; and (iii) conservation of the Hamiltonian via the error $\Delta\mathcal{H}_{N_t} = |\mathcal{H}(y^{N_t}) - \mathcal{H}(Az^{N_t})|$. We refer to [Proposition 4.4](#) for the analytic bound on $\Delta\mathcal{H}_{N_t}$; when AVF is used as numerical time integrator, the error sources in [\(4.15\)](#) that are independent on the reduction take values smaller than 10^{-10} .

In all the following tests, we fix the size of the DEIM basis to $m = 100$ and we let one hyper-parameter vary at a time. Based on the previous test and on [Section 4](#), the sampling strategy in [Algorithm 4.2](#) is adopted for the update of the sampling indices. As a reference, we include the performances of the non-adaptive hyper-reduction, where the offline phase consists of the construction of the DEIM basis from the matrix of the reduced Jacobian at full model snapshots obtained every 20 time steps and 11 training parameters. If the adaptive scheme is employed instead, the offline phase involves solving the full system for $\delta_0 \ll N_t$ time steps and only for the given test parameter. This implies that the offline phase will only require few seconds; on the other hand, the DEIM basis is adapted every δ time steps, which increases the computational cost of the online phase with respect to the non-adaptive scheme.

Effect of the rank r of the DEIM update. We test the hyper-reduced model for different values of the rank r of the DEIM update. Instead of fixing its value we let it change at each update j based on the numerical rank of $S_j^\top R_j C_j^\top$ for a fixed tolerance τ_r . The rank r_j vs. the update number is shown in [Figure 6](#) (left) for different choices of τ_r , while the error $\mathcal{E}_{L^2}^{\text{HR}}$ of the hyper-reduced solution is reported in [Table 2](#). We observe that the error improves as τ_r decreases, while the cost of the online phase of the adaptive hyper-reduction algorithm is almost independent of τ_r . This is due to the fact that the cost of the adaptation procedure is dominated by the computation of the DEIM residual matrix R_j and the SVD of $R_j C_j$ in [Algorithm 4.2](#), whose complexity is $O(dkmw)$, and thus independent of r . Moreover, in this experiment, there is no significant advantage in fixing the tolerance τ_r smaller than 10^{-8} and, correspondingly, performing a DEIM basis update of rank larger than 40.

TABLE 2
Error and computational times of the hyper-reduced model for different rank tolerances τ_r .

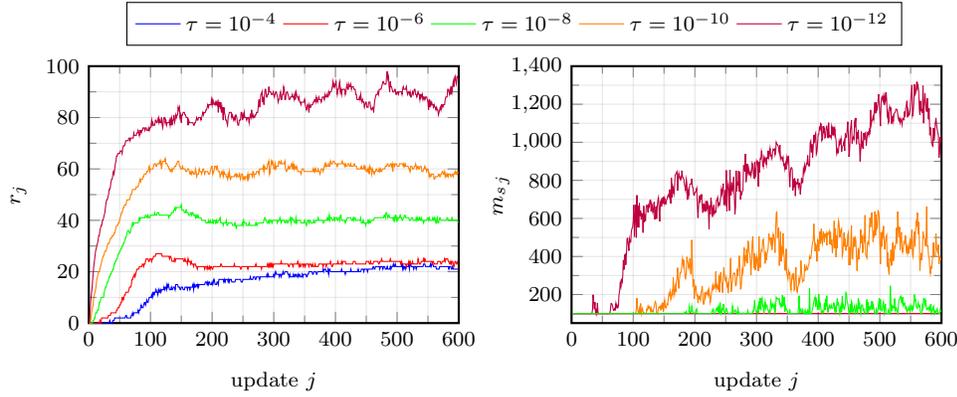
		t_{off}	t_{on}	t_{tot}	$\mathcal{E}_{L^2}^{\text{HR}}$	$\Delta\mathcal{H}_{N_t}$
Non-adaptive		152.77	38.52	191.29	1.04e+00	2.03e+00
$m_s = 250$	$\tau_r = 10^{-4}$	5.44	54.83	60.27	3.04e-03	7.09e-02
	$\tau_r = 10^{-6}$	5.44	55.17	60.61	1.09e-04	2.37e-04
$\delta_0 = \delta = 5$	$\tau_r = 10^{-8}$	5.44	55.30	60.74	6.44e-05	6.88e-04
	$\tau_r = 10^{-10}$	5.44	55.46	60.90	6.17e-05	8.04e-04
$w = 1$	$\tau_r = 10^{-12}$	5.44	55.58	61.02	6.03e-05	7.21e-04

Effect of the number m_s of sampling points. Analogously to the rank of the update, instead of fixing m_s we may select the number of sampling indices based on a tolerance τ_s . More precisely, an index is chosen at the j th adaptation step if the 2-norm of the corresponding row of $R_j C_j$ is larger than τ_s . In this way, the number of sampling indices changes at each adaptation step and it is fixed to be at least equal to m . The results are reported in [Table 3](#). The number of sampling indices selected at each adaptation for different choices of the tolerance is shown in [Figure 6](#) (right). As m_s increases, the error improves while the computational time is almost independent of the value of m_s , and the total time required by the adaptive algorithm is less than one third of the time needed by the non-adaptive algorithm.

TABLE 3

Error and computational times of the hyper-reduced model for different sampling tolerances τ_s .

		t_{off}	t_{on}	t_{tot}	$\mathcal{E}_{L^2}^{\text{HR}}$	$\Delta\mathcal{H}_{N_t}$
Non-adaptive		152.77	38.52	191.29	1.04e+00	2.03e+00
$\delta_0 = \delta = 5$ $w = 1$ $r = \bar{r}$ $\gamma = \delta$	$\tau_s = 10^{-4}$	5.50	53.32	58.82	3.94e-04	1.12e-02
	$\tau_s = 10^{-6}$	5.50	53.81	59.31	3.94e-04	1.12e-02
	$\tau_s = 10^{-8}$	5.50	53.82	59.32	2.86e-04	5.18e-03
	$\tau_s = 10^{-10}$	5.50	54.42	59.92	5.99e-05	9.80e-05
	$\tau_s = 10^{-12}$	5.50	55.49	60.99	4.70e-05	1.52e-04

FIG. 6. NLS. Left: Rank $r_j = \text{rank}(S_j^\top R_j C_j^\top)$ at every update j for different values of $\tau_t = \tau$. Right: Number of sampling indices $m_{s,j}$ at every update j for different values of $\tau_s = \tau$.**Effect of the size w of the time window to evaluate the nonlinear term.**

The role of w in the adaptation is described in Section 4. In this test, we slightly increase the values of δ_0 and δ and set them equal to 10. The reason is that it must hold $w < \delta_0$, and, moreover, it is reasonable to require that $w < \delta$, that is, at each adaptation step, the temporal window only contains snapshots obtained after the last update. In Table 4 we observe that, as expected, the error $\mathcal{E}_{L^2}^{\text{HR}}$ decreases as more information from the past is included in the construction of the DEIM basis update. In parallel, the increase of the window size requires larger computational times in the online phase. However, it can be noticed that the online phase in the case $w = 1$ is only slightly more computationally expensive than the online phase of the non-adaptive algorithm, while the L^2 -error is reduced by almost four orders of magnitude.

TABLE 4

Error and computational times of the hyper-reduced model for different window sizes w .

		t_{off}	t_{on}	t_{tot}	$\mathcal{E}_{L^2}^{\text{HR}}$	$\Delta\mathcal{H}_{N_t}$
Non-adaptive		152.77	38.52	191.29	1.04e+00	2.03e+00
$m_s = 250$ $\delta_0 = \delta = 10$ $r = \bar{r}, \gamma = \delta$	$w = 1$	8.91	45.01	53.92	1.31e-04	2.55e-03
	$w = 2$	8.91	50.82	59.73	1.13e-04	2.67e-03
	$w = 4$	8.91	60.13	69.04	7.44e-05	1.30e-04
	$w = 6$	8.91	69.73	78.64	6.91e-05	1.06e-03

Effect of the number δ of time steps between each DEIM adaptation.

Predictably, the online phase is cheaper for large δ since the adaptations are less frequent, while the error increases, as reported in Table 5.

TABLE 5

Error and computational times of the hyper-reduced model. Effect of the number δ of time steps between each update of the DEIM basis.

	t_{off}	t_{on}	t_{tot}	$\mathcal{E}_{L^2}^{\text{HR}}$	$\Delta\mathcal{H}_{N_t}$
Non-adaptive	152.77	38.52	191.29	1.04e+00	2.03e+00
$m_s = 250$ $\delta = 5$	5.55	55.05	60.60	6.16e-05	5.31e-04
$\delta_0 = 5, w = 1$ $\delta = 10$	5.55	46.40	51.95	1.40e-04	3.08e-03
$r = \bar{r}, \gamma = \delta$ $\delta = 20$	5.55	41.00	46.55	2.50e-03	4.99e-02

Effect of the number of time steps γ between each update of the sampling indices. In view of the discussion on the arithmetic complexity of sampling in [Subsection 4.2](#), it might seem preferable, in terms of computational cost, to update the sampling indices every $\gamma > 1$ time steps. Indeed, the larger γ , the less frequently [Algorithm 4.2](#) is run, including the expensive computation of the full residual. In this test we consider $\gamma = \nu\delta$, i.e. the sampling points are updated every ν DEIM adaptations. We observe in [Table 6](#) that the error of the hyper-reduced solution decreases as γ is halved, with an increase in the runtime of the online phase of less than 5 seconds.

TABLE 6

Error and computational times of the hyper-reduced model. Effect of the number γ of time steps between each update of the sampling indices.

	t_{off}	t_{on}	t_{tot}	$\mathcal{E}_{L^2}^{\text{HR}}$	$\Delta\mathcal{H}_{N_t}$
Non-adaptive	152.77	38.52	191.29	1.04e+00	2.03e+00
$m_s = 250$ $\gamma = 5$	5.51	55.15	60.66	6.16e-05	5.31e-04
$\delta_0 = \delta = 5$ $\gamma = 10$	5.51	50.35	55.86	1.11e-04	2.04e-03
$w = 1$ $\gamma = 20$	5.51	48.20	53.71	8.59e-04	1.83e-02
$r = \bar{r}$ $\gamma = 30$	5.51	47.56	53.07	4.58e-03	9.90e-02

Although the behavior of the adaptive hyper-reduction algorithm under variation of the hyper-parameters is clearly problem-dependent, the numerical tests presented above suggest the following conclusions. Increasing the rank r of the update and the number m_s of sampling points leads to a significant reduction of the error without any major effect on the computational time. On the other hand, the optimal values of r and m_s may depend on the adaptation step j and, hence, it seems preferable to let them vary in time according to a tolerance. The effect of w on the error is less evident: although the error decreases as w increases, the fast growth of the computational time t_{on} outweighs this gain. Finally, small values of δ and γ imply frequent adaptations of both the DEIM basis and sampling indices, and thus reduction in the error, at a mild increase of the computational cost of the simulation.

As a last numerical test, we fix the values of the hyper-parameters and compare the performances of the non-adaptive and adaptive hyper-reduction algorithms as the dimension m of the DEIM space varies, for different choices of the reduced dimension $2k$. We set the hyper-parameters as follows: the tolerance in the choice of the update rank is $\tau_r = 10^{-12}$; the tolerance in the selection of the sampling points is $\tau_s = 10^{-10}$; the temporal window size is $w = 1$; the frequency of the updates is $\delta = \gamma = 5$. In [Figure 7](#) we report the error $\mathcal{E}_{L^2}^{\text{HR}}$ between the hyper-reduced solution and the full model solution vs. m (left plot) and vs. the algorithm runtime (right plot) for $2k = 400$. In [Figure 8](#), we study the computational cost of the hyper-reduction algorithms for two more dimensions of the reduced space, namely for $2k \in \{160, 280\}$. The reported times are computed as averages over 5 runs.

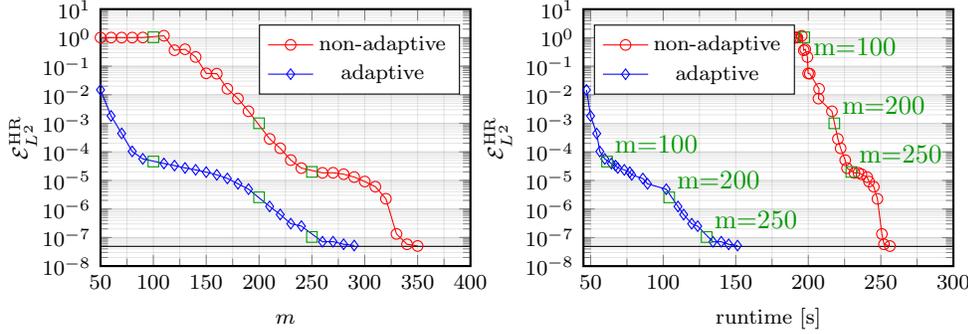


FIG. 7. NLS. Reduced model of size $2k = 400$. Left: error $\mathcal{E}_{L^2}^{\text{HR}}$ vs. DEIM size m . Right: $\mathcal{E}_{L^2}^{\text{HR}}$ vs. total algorithm runtime for different values of m . The black line represents the error between the reduced solution and the full solution. The computational time to solve the reduced model is 230 s. The computational time of the offline phase of the hyper-reduction is 152.8 s in the non-adaptive case, 5.5 s in the adaptive case.

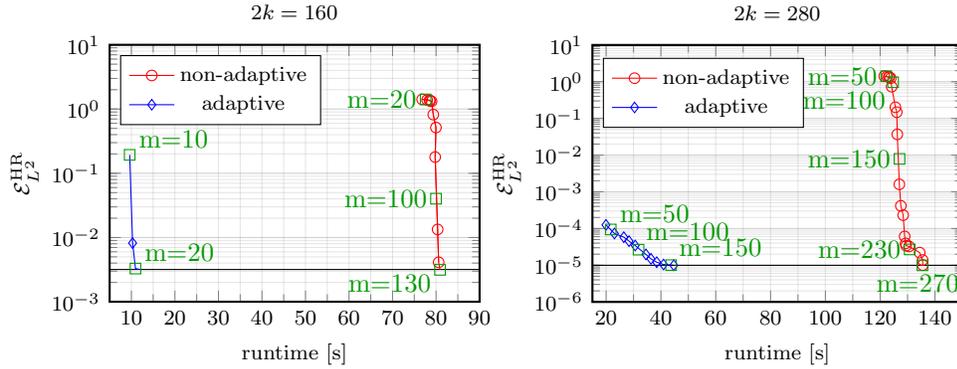


FIG. 8. NLS. Hyper-reduction error $\mathcal{E}_{L^2}^{\text{HR}}$ vs. total algorithm runtime for different values of m . Reduced model of size $2k = 160$ (left) and $2k = 280$ (right). For $2k = 160$: the computational time to solve the reduced model is 66 s; the offline phase of the hyper-reduction takes 69.7 s in the non-adaptive case and 0.5 s in the adaptive case. For $2k = 280$: the computational time to solve the reduced model is 126 s; the offline phase of the hyper-reduction takes 109.6 s in the non-adaptive case and 1.2 s in the adaptive case.

Figures 7 and 8 show that, to achieve a given error, the adaptive scheme allows a smaller DEIM basis compared to the non-adaptive one. This translates in a reduced computational cost, despite the fact that the adaptive algorithm requires extra online operations to perform the DEIM update. This gain in the total runtime is also due to the fact that a more expensive online phase for adaptation is compensated by a cheaper offline phase compared to the non-adaptive strategy. As an example, the total runtime is reduced by a factor 3 for $m = 100$ when $2k = 400$, and by a factor 4 when $2k = 280$. Moreover, for a given DEIM size m , the adaptive scheme yields smaller errors than the non-adaptive one: for example, for $m = 200$, the error in the adaptive case is almost three orders of magnitude lower when $2k = 400$, and a similar gain is observed for $m = 150$ when $2k = 280$. Finally, when the dimension of the reduced space is $2k = 400$, the hyper-reduced model achieves the error of the reduced model for $m = 290$ in the adaptive case, and for $m = 350$ in the non-adaptive one; for these values of m , the adaptive algorithm requires about 40% less runtime

than the non-adaptive scheme. This gain in computational efficiency becomes more evident when the reduced space is smaller. For $2k = 280$, the DEIM modes required to achieve the reference error in the adaptive case are about half of those needed in the non-adaptive case (160 vs. 270), with an approximately 70% saving of the total computational time. With an even smaller reduced space, $2k = 160$, the adaptive hyper-reduced model converges to the reduced model error for $m = 25$, compared to $m = 130$ in the non-adaptive case, and the total runtime is about 85% smaller; even comparing the online phases alone, solving the adaptive scheme with $m = 20$ is about as cheap as solving the non-adaptive scheme for $m = 80$, while the error is over 2 orders of magnitude smaller. Finally, these numerical experiments show that the beneficial properties of the proposed adaptive strategy become more apparent when the hyper-reduction is performed on a relatively small reduced space. This is related to the decay of the singular values of the reduced Jacobian highlighted in [Figure 4](#).

6. Concluding remarks. We have proposed a hyper-reduction method to deal with large-scale parametric dynamical systems characterized by nonlinear gradient fields. Its combination with symplectic reduced basis methods enables efficient numerical simulations of Hamiltonian systems with general nonlinearities. Indeed the resulting hyper-reduced models can be solved at a computational cost independent of the size of the full model and without compromising the physical properties of the dynamics thanks to the preservation of the gradient structure of the velocity field. Although we have focused on Hamiltonian dynamical systems, the method can be applied to any gradient field, in e.g. gradient flows or in port-Hamiltonian problems.

Some questions remain open. A rigorous connection between the reducibility properties of the solution space, of the space of nonlinear gradient fields and of the space of reduced Jacobian matrices is not understood yet. This could provide useful insights on when the simulation of a system can significantly benefit from hyper-reduction and whether an adaptive approach is to be preferred over a global one.

Moreover, our hyper-reduction strategy hinges on the properties of the reduced space. We observe that a larger reduced space improves the approximation of the state but demands a larger DEIM basis for accurate hyper-reduction. A natural idea to address this shortcoming is to consider symplectic reduced spaces that can evolve in time, as done e.g. in [\[22\]](#). This would allow us to deal with small reduced spaces and hence highly reducible Jacobian matrices (i.e. small DEIM bases as well). The combination of evolving reduced spaces with adaptive hyper-reduction provides an interesting possible direction of future investigation.

Acknowledgments. CP would like to thank Benjamin Peherstorfer for interesting discussions on adaptive DEIM preliminary to this work.

REFERENCES

- [1] B. M. AFKHAM AND J. S. HESTHAVEN, *Structure preserving model reduction of parametric Hamiltonian systems*, SIAM J. Sci. Comput., 39 (2017), pp. A2616–A2644.
- [2] M. BARRAULT, Y. MADAY, N. C. NGUYEN, AND A. T. PATERA, *An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations*, C. R. Math. Acad. Sci. Paris, 339 (2004), pp. 667–672.
- [3] P. BUCHFINK, A. BHATT, AND B. HAASDONK, *Symplectic model order reduction with non-orthonormal bases*, Math. Comput. Appl., 24 (2019), pp. Paper No. 43, 26.
- [4] P. BUCHFINK, B. HAASDONK, AND S. RAVE, *PSD-Greedy basis generation for structure-preserving model order reduction of Hamiltonian systems*, Proceedings of the Conference Algoritmy, (2020), pp. 151–160.

- [5] A. CANNAS DA SILVA, *Lectures on symplectic geometry*, vol. 1764 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2001.
- [6] K. CARLBERG, R. TUMINARO, AND P. BOGGS, *Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics*, SIAM J. Sci. Comput., 37 (2015), pp. B153–B184.
- [7] E. CELLEDONI, V. GRIMM, R. I. MCLACHLAN, D. I. MCLAREN, D. O’NEALE, B. OWREN, AND G. R. W. QUISPTEL, *Preserving energy resp. dissipation in numerical PDEs using the “average vector field” method*, J. Comput. Phys., 231 (2012), pp. 6770–6789.
- [8] S. CHATURANTABUT, C. BEATTIE, AND S. GUGERCIN, *Structure-preserving model reduction for nonlinear port-Hamiltonian systems*, SIAM J. Sci. Comput., 38 (2016), pp. B837–B865.
- [9] S. CHATURANTABUT AND D. C. SORENSEN, *Nonlinear model reduction via discrete empirical interpolation*, SIAM J. Sci. Comput., 32 (2010), pp. 2737–2764.
- [10] S. CHATURANTABUT AND D. C. SORENSEN, *A state space error estimate for POD-DEIM nonlinear model reduction*, SIAM J. Numer. Anal., 50 (2012), pp. 46–63.
- [11] 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.
- [12] C. FARHAT, T. CHAPMAN, AND P. AVERY, *Structure-preserving, stability, and accuracy properties of the energy-conserving sampling and weighting method for the hyper reduction of nonlinear finite element dynamic models*, Internat. J. Numer. Methods Engrg., 102 (2015), pp. 1077–1110.
- [13] Y. GONG, Q. WANG, AND Z. WANG, *Structure-preserving Galerkin POD reduced-order modeling of Hamiltonian systems*, Comput. Methods Appl. Mech. Engrg., 315 (2017), pp. 780–798.
- [14] T. H. GRONWALL, *Note on the derivatives with respect to a parameter of the solutions of a system of differential equations*, Ann. of Math. (2), 20 (1919), pp. 292–296.
- [15] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric numerical integration*, vol. 31 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 2006.
- [16] N. HALKO, P. G. MARTINSSON, AND J. A. TROPP, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Review, 53 (2011), pp. 217–288.
- [17] J. S. HESTHAVEN AND C. PAGLIANTINI, *Structure-preserving reduced basis methods for Poisson systems*, Math. Comp., 90 (2021), pp. 1701–1740.
- [18] J. S. HESTHAVEN, C. PAGLIANTINI, AND N. RIPAMONTI, *Adaptive symplectic model order reduction of parametric particle-based Vlasov-Poisson equation*, 2022, <https://arxiv.org/abs/2201.05555>.
- [19] J. S. HESTHAVEN, C. PAGLIANTINI, AND G. ROZZA, *Reduced basis methods for time-dependent problems*, Acta Numerica, 31 (2022), p. 265–345.
- [20] S. LALL, P. KRYSL, AND J. E. MARSDEN, *Structure-preserving model reduction for mechanical systems*, Phys. D, 184 (2003), pp. 304–318.
- [21] J. E. MARSDEN AND T. S. RATIU, *Introduction to mechanics and symmetry*, vol. 17 of Texts in Applied Mathematics, Springer-Verlag, New York, second ed., 1999.
- [22] C. PAGLIANTINI, *Dynamical reduced basis methods for Hamiltonian systems*, Numer. Math., 148 (2021), pp. 409–448.
- [23] B. PEHERSTORFER, *Model reduction for transport-dominated problems via online adaptive bases and adaptive sampling*, SIAM J. Sci. Comput., 42 (2020), pp. A2803–A2836.
- [24] 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.
- [25] L. PENG AND K. MOHSENI, *Symplectic model reduction of Hamiltonian systems*, SIAM J. Sci. Comput., 38 (2016), pp. A1–A27.
- [26] G. R. W. QUISPTEL AND D. I. MCLAREN, *A new class of energy-preserving numerical integration methods*, J. Phys. A, 41 (2008), pp. 045206, 7.
- [27] H. SHARMA, Z. WANG, AND B. KRAMER, *Hamiltonian operator inference: Physics-preserving learning of reduced-order models for Hamiltonian systems*, Physica D: Nonlinear Phenomena, 431 (2022), p. 133122.
- [28] S. SULTANA AND Z. RAHMAN, *Hamiltonian formulation for water wave equation*, Open Journal of Fluid Dynamics, 3 (2013), pp. 75–81.
- [29] M. UZUNCA, B. KARASÖZEN, AND S. YILDIZ, *Structure-preserving reduced-order modeling of Korteweg-de Vries equation*, Mathematics and Computers in Simulation, 188 (2021), pp. 193–211.
- [30] Z. WANG, *Structure-preserving Galerkin POD-DEIM reduced-order modeling of Hamiltonian systems*, 2021, <https://arxiv.org/abs/2103.00388>.