ELSEVIER

Energy-conserving hyper-reduction and temporal localization for reduced order models of the incompressible Navier-Stokes equations

R.B. Klein [a,b,*], B. Sanderse [a]

[a] *Centrum Wiskunde & Informatica, Science Park 123, Amsterdam, the Netherlands*
[b] *Delft University of Technology, Process and Energy, Leeghwaterstraat 39, Delft, the Netherlands*

ABSTRACT

A novel hyper-reduction method is proposed that conserves kinetic energy and momentum for reduced order models of the incompressible Navier-Stokes equations. The main advantage of conservation of kinetic energy is that it endows the hyper-reduced order model (hROM) with a nonlinear stability property. The new method poses the discrete empirical interpolation method (DEIM) as a minimization problem and subsequently imposes constraints to conserve kinetic energy. Two methods are proposed to improve the robustness of the new method against error accumulation: oversampling and Mahalanobis regularization. Mahalanobis regularization has the benefit of not requiring additional measurement points. Furthermore, a novel method is proposed to perform energy- and momentum-conserving temporal localization with the principle interval decomposition: new interface conditions are derived such that energy and momentum are conserved for a full time-integration instead of only during separate intervals. The performance of the new energy- and momentum-conserving hyper-reduction methods and the energy- and momentum-conserving temporal localization method is analysed using three convection-dominated test cases; a shear-layer roll-up, two-dimensional homogeneous isotropic turbulence and a time-periodic inviscid flow consisting of a vortex in a uniform background flow. Our main finding is that energy conservation in combination with oversampling or regularization leads to a robust method with excellent long time stability properties. When any of these two ingredients is missing, accuracy and/or stability is significantly impaired.

1. Introduction

Computational fluid dynamics (CFD) has become an integral part of many modern engineering applications. The increase in computational power in recent decades has allowed engineers to model increasingly larger fluid dynamical systems. However, many modern applications are of a multi-query or real-time nature e.g. design optimization [82] and uncertainty quantification [29] or real-time control [81,71] and digital twin technology [43]. These applications still pose prohibitively large computational costs. Reduced order models (ROMs) have been proposed as a solution to this problem. A ROM is a type of surrogate model that approximates the high-dimensional full scale model, or full order model (FOM), in a low-dimensional way by finding approximate formulations of

involved quantities or operators. The low-dimensionality of the model consequently makes the ROM significantly cheaper to evaluate than the FOM.

Traditionally in the CFD community, these ROMs have been constructed by projecting the fluid dynamics equations of interest onto low-dimensional linear spaces obtained from the proper orthogonal decomposition (POD) algorithm [47,93], using either Galerkin [87,81,95,50,64] or Petrov-Galerkin [18,38,17] projection. More recently, alternatives have also been explored. ROMs have been constructed without availability of a FOM by inferring the ROM from data using operator inference methods [76,92,12,63]. Machine learning methods like convolutional autoencoders have been used in a projective sense [57,80] and also inference methods [69,62] have been applied to obtain nonlinear low-dimensional approximations. Other nonlinear dimensionality reduction methods like quadratic manifolds [10,34] and diffusion maps [94] have also been leveraged. However, the traditional POD-Galerkin methods remain powerful and this article will primarily deal with these methods and their natural extensions like the principal interval decomposition (PID) [14,49].

Nevertheless, the traditional methods possess several limitations. For turbulent (and convection-dominated) flows that are of engineering interest it is well-known that linear, projection-based ROMs suffer from stability and accuracy issues [87,33,8,5,64]. Efforts have been made to solve this issue, an overview is provided by [33,87]. A promising solution is structure-preservation, this entails constructing ROMs such that the underlying physics of fluid flows are respected. Especially conservation of kinetic energy is an important physical principle to uphold in a ROM with regards to stability as it bounds the norm of the solution [87,2,21,96].

Current attempts at constructing energy-conserving ROMs of fluid flow can be classified into four categories. The first category is constrained optimization projection [19,91,39,13,48]; here model reduction is cast into an optimization problem and constrained to preserve structure. A second category is formed by symmetry-preservation [87,2,21,96] where symmetries of the continuous analogues of ROM operators have been preserved resulting in the conservation of energy or entropy and associated stability. In [36,23,78,46,1,77,45] Hamiltonian physics are preserved in a low-dimensional setting, forming a third category. Finally, in [67,99, 59,68,9,4] physics-informed data-driven approaches using inference and machine learning methods have been adopted establishing a fourth category.

The idea of structure preservation is elegant but does not resolve directly the well-known issue that reducing nonlinear models using POD requires intermediate lifting of the reduced representation to the high-dimensional (FOM) spaces. Thus, although the ROM is low-dimensional, the computational effort to evaluate it is still high-dimensional, defeating its purpose. Methods to overcome this problem are referred to as hyper-reduction methods [16]. For sufficiently simple equations exact hyper-reduction methods exist that eliminate computational dependence on the FOM dimensions [87,3]. This can be done when the underlying nonlinearity is of polynomial nature. Yet, these methods can become prohibitively expensive for larger ROMs as noted in [87]. Hence, approximate hyper-reduction methods may be considered since these generally have better scaling properties. Examples of such approximate hyper-reduction methods are the empirical interpolation method (EIM) [11,37] and its discrete counterpart the discrete empirical interpolation method (DEIM) [24,26,25,11], Gauss-Newton with approximated tensors (GNAT) [20,18], energy-conserving sampling and weighting (ECSW) [38,32,31] and missing point estimation (MPE) [7].

Many existing hyper-reduction methods do not preserve the structure of the operators to which they are applied and thus stability can be lost. The field of model reduction of Hamiltonian systems has offered some solutions for this; [23] proposes a Hamiltonian-conserving DEIM variant for systems with Hamiltonian functionals with non-quadratic terms; [97] improves this DEIM variant and [66] proposes a Hamiltonian-conserving DEIM variant for nonlinear Hamiltonian operators that preserves skew-symmetry. An approach similar to [97] was proposed in [73]. However, the incompressible Navier-Stokes equations considered in this work have a quadratic Hamiltonian functional in the inviscid case [72,6] (kinetic energy) making the first two methods ([23,97]) inapplicable. Furthermore, the approximate method proposed in [66] scales computationally the same as the prohibitively expensive exact method used in [87].

In this article we propose a novel energy- and momentum-conserving hyper-reduction method. The method is similar to the DEIM and gappy-POD [30] in which FOM operators are projected on low-dimensional spaces based on a small (carefully chosen) subset of spatial operator evaluations referred to as measurements. The new method is capable of robustly, accurately and efficiently dealing with convection-dominated flows described by the incompressible Navier-Stokes equations. The main idea will be to relax the DEIM interpolation requirement and use the resulting available degree of freedom to enforce energy conservation. Furthermore, we will propose two methods to further enhance the robustness of the energy- and momentum-conserving hyper-reduction method. Namely, oversampling [75,83], resembling the gappy-POD approach [30], and the new approach of regularization using the Mahalanobis distance (which, to our knowledge, has not yet been applied to the DEIM or gappy-POD). Finally, we will extend the feasibility of the hyper-reduced order model (hROM) with a newly proposed energy- and momentum-conserving localization method based on the PID. Though the methods proposed in this work pertain to the way the operators are projected based on some measurements - which is not specific to either DEIM or gappy-POD - we will mostly refer to them as variants of the DEIM. This could be considered inaccurate as the DEIM is often characterized by the way the measurement points are chosen [24]. In theory our methods can also be applied in the more general context of gappy-POD. However, our work has only been carried out and tested with DEIM-based measurement point selection procedures, hence we prefer to limit our naming to the DEIM.

This article will be organized as follows. First, the governing equations will be introduced in section 2, together with an energy- and momentum-conserving FOM and ROM. Following this, the energy- and momentum-conserving hyper-reduction method will be proposed in section 3. In section 4 the new energy- and momentum-conserving interval decomposition approach for convection-dominated flows will be discussed. Finally, in section 5 the performance of the proposed energy- and momentum-conserving hyper-reduction and temporal localization methods will be analysed using three convection-dominated test cases.

2. Preliminaries: Navier-Stokes equations, FOM and ROM, energy equations

2.1. Incompressible Navier-Stokes equations and kinetic energy conservation

In this work we consider the incompressible Navier-Stokes equations:

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) = -\nabla p + \nu \Delta \boldsymbol{u}, \tag{1}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{2}$$

on a domain $\Omega \subset \mathbb{R}^d$, where $d \in \{2, 3\}$, with periodic boundary conditions. We denote with $\mathbb{R}^+ \subset \mathbb{R}$ the set of positive real numbers including 0. In equations (1)-(2) $\boldsymbol{u} : \Omega \times \mathbb{R}^+ \to \mathbb{R}^d$ is the velocity field, $p : \Omega \times \mathbb{R}^+ \to \mathbb{R}$ the modified pressure field (pressure scaled by density), $\boldsymbol{x} \in \Omega$, $t \in \mathbb{R}^+$ denotes time and $\nu \in \mathbb{R}^+$ is the kinematic viscosity. Several quantities related to the flow will also be considered, namely kinetic energy $K : \mathbb{R}^+ \to \mathbb{R}^+$ and momentum $\boldsymbol{P} : \mathbb{R}^+ \to \mathbb{R}^d$. Kinetic energy is defined as:

$$K(t) := \frac{1}{2} \|\boldsymbol{u}\|_{L^2}^2, \tag{3}$$

with $\boldsymbol{u}$ evaluated at t and where:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{L^2} := \int_\Omega \boldsymbol{u} \cdot \boldsymbol{v} \, d\Omega, \quad \|\boldsymbol{u}\|_{L^2} := \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle_{L^2}},$$

are the L^2-inner product and induced L^2-norm, respectively. Momentum is defined as:

$$\boldsymbol{P}(t) := \int_\Omega \boldsymbol{u} \, d\Omega. \tag{4}$$

Denoting $C(\boldsymbol{u}, \boldsymbol{u}) := \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u})$, $\mathcal{D}\boldsymbol{u} := \Delta \boldsymbol{u}$, $\mathcal{G}p := \nabla p$ and $\mathcal{W}\boldsymbol{u} := \nabla \cdot \boldsymbol{u}$, an evolution equation of the total kinetic energy can be found by differentiating (3) and substituting (1):

$$\frac{dK}{dt} = -\langle \boldsymbol{u}, C(\boldsymbol{u}, \boldsymbol{u}) \rangle_{L^2} - \langle \boldsymbol{u}, \mathcal{G}p \rangle_{L^2} + \nu \langle \boldsymbol{u}, \mathcal{D}\boldsymbol{u} \rangle_{L^2}.$$

This evolution equation can be further simplified when the following is considered. Since the convection operator is skew-adjoint given condition (2), it holds that $\langle \boldsymbol{u}, C(\boldsymbol{u}, \boldsymbol{u}) \rangle_{L^2} = -\langle C(\boldsymbol{u}, \boldsymbol{u}), \boldsymbol{u} \rangle_{L^2} = -\langle \boldsymbol{u}, C(\boldsymbol{u}, \boldsymbol{u}) \rangle_{L^2}$, meaning that $\langle \boldsymbol{u}, C(\boldsymbol{u}, \boldsymbol{u}) \rangle_{L^2} = 0$. Additionally, since the gradient operator and the divergence operator are each other's negated adjoints it can be stated that $\langle \boldsymbol{u}, \mathcal{G}p \rangle_{L^2} = -\langle \mathcal{W}\boldsymbol{u}, p \rangle_{L^2} = 0$ due to equation (2). Finally, the negative-definiteness of the diffusion operator can be employed to simplify the evolution equation of the total kinetic energy further to:

$$\frac{dK}{dt} = -\nu \|\nabla \boldsymbol{u}\|_{L^2}^2 \leq 0. \tag{5}$$

Equation (5) implies that the kinetic energy, or the norm of the velocity field $\boldsymbol{u}$, is a monotonically decreasing quantity for periodic or homogeneous Dirichlet boundary conditions and is conserved in the inviscid case ($\nu = 0$). It will become clear soon that mimicking this property discretely will be crucial to nonlinear stability of the FOM and ROMs.

Furthermore, it follows from integrating equation (1) over the domain Ω that:

$$\frac{d\boldsymbol{P}}{dt} = 0, \tag{6}$$

for periodic boundary conditions. It is now our goal to *mimic these conservation properties at the discrete and especially the reduced level*.

2.2. Energy- and momentum-conserving FOM

Equation (1) is discretized with a finite volume method (FVM) on a staggered grid, resulting in a system of coupled ordinary differential equations (ODEs) complemented by a set of linear constraints:

$$\Omega_h \frac{d\boldsymbol{u}_h}{dt} + C_h(\boldsymbol{u}_h) = -G_h \boldsymbol{p}_h + \nu D_h \boldsymbol{u}_h, \tag{7}$$

$$W_h \boldsymbol{u}_h = 0. \tag{8}$$

Here $\boldsymbol{u}_h : \mathbb{R}^+ \to \mathbb{R}^N$ are the numerical velocity values arranged in a vector, $\Omega_h \in \mathbb{R}^{N \times N}$ is a diagonal matrix containing finite volume sizes, $C_h : \mathbb{R}^N \to \mathbb{R}^N$ is the spatial discretization of the nonlinear convection operator, $G_h : \mathbb{R}^{N_p} \to \mathbb{R}^N$ is the spatial discretization of the gradient operator, $\boldsymbol{p}_h : \mathbb{R}^+ \to \mathbb{R}^{N_p}$ are the numerical pressure values arranged in a vector, $D_h : \mathbb{R}^N \to \mathbb{R}^N$ is the spatial discretization of the diffusion operator, $W_h : \mathbb{R}^N \to \mathbb{R}^{N_p}$ is the spatial discretization of the divergence operator, $N = N_u + N_v$ is the total number of velocity unknowns and N_u and N_v are the numbers of velocity unknowns in the x and y directions respectively (for the case of two-dimensional domains) and N_p is the number of pressure unknowns. Relations (7)-(8) are in turn complemented by a vector of initial conditions $\boldsymbol{u}_0 = \boldsymbol{u}_h(0)$ and suitable boundary conditions. The choice of boundary conditions for the flows considered in this work are periodic.

We use the discretization as described in [85,87] such that the following three important properties of the continuous operators are inherited by their spatial discretizations. First, the convection operator, which can be written in quasi-linear form as $C_h(\boldsymbol{u}_h) = \widetilde{C}_h(\boldsymbol{u}_h)\boldsymbol{u}_h$ (see [87] for details), is skew-symmetric:

$$\widetilde{C}_h(\boldsymbol{u}_h) = -\widetilde{C}_h(\boldsymbol{u}_h)^T. \tag{9}$$

The quasi-linear form is not essential for the development of our hyper-reduction methods, but makes the proof of energy conservation straightforward. Second, the duality between gradient and divergence operators holds:

$$G_h = -W_h^T. \tag{10}$$

Third, the diffusive operator is symmetric negative-definite, allowing the general decomposition (e.g., a Cholesky decomposition):

$$D_h = -Q_h^T Q_h. \tag{11}$$

We remark that the conservation properties discussed in this article are not specific to the method described in [85,87], any other finite volume discretization of (1)-(2) satisfying the above properties can also be used.

Using these properties and the telescoping property of the FVM, discretization (7)-(8) conserves discrete analogues of mass, total momentum and total kinetic energy. To show this, we first define discrete total kinetic energy and momentum on the basis of an inner product. To this end, the Ω_h-inner product is introduced. The Ω_h-inner product and its induced norm are defined by:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\Omega_h} := \langle \boldsymbol{u}, \Omega_h \boldsymbol{v} \rangle, \quad ||\boldsymbol{u}||_{\Omega_h} := \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle_{\Omega_h}},$$

where $\langle \cdot, \cdot \rangle$, without subscript, denotes the standard Euclidean inner product. After deriving an evolution equation for the discrete total kinetic energy it will be clear why this is a natural choice. Using the Ω_h-inner product, the discrete total kinetic energy $K_h : \mathbb{R}^+ \to \mathbb{R}^+$ is defined as:

$$K_h(t) := \frac{1}{2} ||\boldsymbol{u}_h(t)||_{\Omega_h}^2.$$

The discrete total momentum $\boldsymbol{P}_h : \mathbb{R}^+ \to \mathbb{R}^d$ is defined as:

$$\left(\boldsymbol{P}_h(t) \right)_i := \langle \boldsymbol{e}_i, \boldsymbol{u}_h(t) \rangle_{\Omega_h},$$

where $\boldsymbol{e}_i \in \mathbb{R}^N$ is a vector of ones at the vector indices where $\boldsymbol{u}_h(t)$ stores velocity components in direction $i \in \{1, ..., d\}$ and zeros elsewhere.

An evolution equation for discrete kinetic energy is found by temporal differentiation of $K_h(t)$:

$$\frac{dK_h}{dt} = \frac{1}{2}\frac{d}{dt}\langle \boldsymbol{u}_h, \boldsymbol{u}_h \rangle_{\Omega_h} = -\left\langle \boldsymbol{u}_h, \widetilde{C}_h(\boldsymbol{u}_h)\boldsymbol{u}_h \right\rangle - \langle \boldsymbol{u}_h, G_h p_h \rangle + \nu \langle \boldsymbol{u}_h, D_h \boldsymbol{u}_h \rangle$$

$$= -\nu ||Q_h \boldsymbol{u}_h||^2 \leq 0, \tag{12}$$

where $|| \cdot ||$, without subscript, denotes the Euclidean norm induced by the Euclidean inner product. The simplification in the second line follows from the application of properties (9)-(11). Indeed, evolution equation (12) is a discrete analogue to equation (5) and shows that $K_h(t)$, and so the norm of $\boldsymbol{u}_h(t)$, is a monotonically decreasing quantity. The solution of the semi-discrete system of ODEs (7) can thus be bounded from above by:

$$||\boldsymbol{u}_h(t)||_{\Omega_h} \leq ||\boldsymbol{u}_0||_{\Omega_h},$$

guaranteeing stability of $\boldsymbol{u}_h(t)$.

It is also a well-known property of the FVM, referred to as the telescoping property [98], that:

$$\left(\frac{d\boldsymbol{P}_h}{dt} \right)_i = \frac{d}{dt}\langle \boldsymbol{e}_i, \boldsymbol{u}_h \rangle_{\Omega_h} = -\left\langle \boldsymbol{e}_i, \widetilde{C}_h(\boldsymbol{u}_h)\boldsymbol{u}_h \right\rangle - \langle \boldsymbol{e}_i, G_h p_h \rangle + \nu \langle \boldsymbol{e}_i, D_h \boldsymbol{u}_h \rangle$$

$$= 0, \quad \forall i \in \{1, ..., d\}. \tag{13}$$

This constitutes a discrete analogue to (6) and shows that discrete total momentum is also conserved.

2.3. Energy- and momentum-conserving ROM

To construct a reduced order model as in [87] we make the assumption that for any $t \in [0, T]$ there are elements $\boldsymbol{u}_r(t)$ of a low-dimensional linear subspace $\mathcal{V} \subset \mathbb{R}^N$ that accurately approximate $\boldsymbol{u}_h(t)$. Here, we denote $r := \dim(\mathcal{V})$ and the low dimensionality of $\mathcal{V}$ implies that $r \ll N$. Furthermore, we consider $\mathcal{V} \subset \ker(W_h)$ such that all $\boldsymbol{u}_r(t)$ automatically satisfy condition (8). Let $\Phi \in \mathbb{R}^{N \times r}$ be a POD basis for $\mathcal{V}$ that is orthonormal in the Ω_h-inner product, i.e. $\left\langle \Phi_{,i}, \Phi_{,j} \right\rangle_{\Omega_h} = \delta_{ij}$,[1] then we write:

---

[1] We use commas in the notation $(A_{,j})$ to the denote the jth column of matrix A, similarly $(A_{i,})$ denotes the ith row.

$$\boldsymbol{u}_h(t) \approx \boldsymbol{u}_r(t) := \Phi \boldsymbol{a}(t) \in \mathcal{V}. \tag{14}$$

Here $\boldsymbol{a} : \mathbb{R}^+ \to \mathbb{R}^r$ are the generalized coordinates in $\mathcal{V}$ associated to the POD basis Φ. Substituting this approximation into the FOM (8) we obtain an overdetermined system. To obtain a solvable system, we test the overdetermined system against the POD modes in Φ and impose the Galerkin condition. This results in a pressure-free[2] ROM:

$$\frac{d\boldsymbol{a}}{dt} = -\widetilde{C}_r(\boldsymbol{a})\boldsymbol{a} + \nu D_r \boldsymbol{a}, \tag{15}$$

where $\widetilde{C}_r(\boldsymbol{a}) := \Phi^T \widetilde{C}_h(\Phi \boldsymbol{a})\Phi$ and $D_r := \Phi^T D_h \Phi$.

In order to consider energy and momentum conservation we first define the following quantities. A reduced order representation of kinetic energy is defined as:

$$K_r(t) := \frac{1}{2} ||\boldsymbol{u}_r(t)||^2_{\Omega_h} = \frac{1}{2} ||\boldsymbol{a}(t)||^2. \tag{16}$$

For total momentum a reduced order representation is:

$$\left(\boldsymbol{P}_r(t)\right)_i := \langle \boldsymbol{e}_i, \boldsymbol{u}_r(t) \rangle_{\Omega_h} = \left\langle \Phi^T \Omega_h \boldsymbol{e}_i, \boldsymbol{a}(t) \right\rangle.$$

Noting that symmetry properties (9) and (11) are still satisfied by the reduced operators in (15), the evolution equation for $K_r(t)$ follows by differentiating (16):

$$\begin{aligned}
\frac{dK_r}{dt} &= \frac{1}{2}\frac{d}{dt}\langle \boldsymbol{a}, \boldsymbol{a} \rangle = -\left\langle \boldsymbol{a}, \widetilde{C}_r(\boldsymbol{a})\boldsymbol{a} \right\rangle + \nu \langle \boldsymbol{a}, D_r \boldsymbol{a} \rangle \\
&= -\nu ||Q_r \boldsymbol{a}||^2 \le 0,
\end{aligned} \tag{17}$$

where $Q_r := Q_h \Phi$. This equation is the ROM analogue of equation (5) and shows that the norm of the generalized coordinates satisfies:

$$||\boldsymbol{a}(t)|| \le ||\boldsymbol{a}_0||,$$

with equality for $\nu = 0$.

To conserve total momentum at the reduced level we require $\boldsymbol{e}_i \in \mathcal{V} \ \forall i \in \{1, ..., d\}$. In short, this is achieved by applying the SVD to an adapted snapshot matrix, in which the projection of the snapshot data on the $\boldsymbol{e}_i$'s is subtracted from the snapshots, and the resulting POD modes are subsequently enriched with the $\boldsymbol{e}_i$'s to form the final basis. For more details, see [87]. An evolution equation for $\boldsymbol{P}_r(t)$ is then found as:

$$\begin{aligned}
\left(\frac{d\boldsymbol{P}_r}{dt}\right)_i &= \frac{d}{dt}\left\langle \Phi^T \Omega_h \boldsymbol{e}_i, \boldsymbol{a} \right\rangle = -\left\langle \Phi^T \Omega_h \boldsymbol{e}_i, \widetilde{C}_r(\boldsymbol{a})\boldsymbol{a} \right\rangle + \nu \left\langle \Phi^T \Omega_h \boldsymbol{e}_i, D_r \boldsymbol{a} \right\rangle \\
&= -\left\langle \boldsymbol{e}_i, \widetilde{C}_h(\Phi \boldsymbol{a})\Phi \boldsymbol{a} \right\rangle + \nu \left\langle \boldsymbol{e}_i, D_h \Phi \boldsymbol{a} \right\rangle \\
&= 0.
\end{aligned} \tag{18}$$

In the second line the telescoping property of the FOM operators was evoked.

Finally, a POD basis Φ satisfying all required criteria (span(Φ) $\subset$ ker(W_h), Ω_h-orthonormality and containing all $\boldsymbol{e}_i/||\boldsymbol{e}_i||_{\Omega_h}$ in its columns) is constructed from snapshot data using an altered POD method as described in [87]. The FOM and the ROM can both be integrated in time with energy-conserving Runge-Kutta methods to conserve energy fully discretely (see [87,86] for further details).

3. An energy- and momentum-conserving discrete empirical interpolation method

3.1. Introduction to DEIM

The ROM (15) as constructed in the previous section is r-dimensional, but evaluating the convective terms in a naive fashion still requires a computational effort that scales with the FOM dimension N. One approach to alleviate this issue is the exact tensor decomposition. This method results in an exact low-dimensional representation of the convection operator but scales with $\mathcal{O}(r^3)$, which becomes computationally expensive for realistic values of r [87]. A more efficient alternative approach, which will be considered here, is the discrete empirical interpolation method (DEIM) [24]. Using the DEIM we will construct a so-called hyper-reduced order model (hROM).

The DEIM approximates $C_h(\boldsymbol{u})$ with elements from a low-dimensional linear subspace $\mathcal{M} \subset \mathbb{R}^N$. Here, we denote $m := \dim(\mathcal{M})$ and the low dimensionality of $\mathcal{M}$ implies that $m \ll N$. Let $M \in \mathbb{R}^{N \times m}$ be an orthonormal basis for $\mathcal{M}$ constructed from snapshot data of the operator using the POD algorithm, where the snapshot data is gathered in a matrix $\Xi \in \mathbb{R}^{N \times n_s}$ given as:

$$\Xi = \left[C_h(\boldsymbol{u}_h(t^0)), C_h(\boldsymbol{u}_h(t^0 + \Delta t)), ..., C_h(\boldsymbol{u}_h(t^0 + (n_s - 1)\Delta t)) \right]. \tag{19}$$

---

2 Note that since $\mathcal{V} \subset$ ker(W_h), we can write $\Phi^T G_h p_h = -(W_h \Phi)^T p_h = 0$, making the ROM velocity-only.

The basis M thus satisfies $\langle M_{,i}, M_{,j} \rangle = \delta_{ij}$. The DEIM approximation then takes the form:

$$C_h(\Phi a) \approx M c \in \mathcal{M}, \tag{20}$$

where $c \in \mathbb{R}^m$ are the generalized coordinates of the approximation in $\mathcal{M}$ associated to basis M and are referred to as DEIM coordinates. The linear nature of the approximation allows the Galerkin projection $\Phi^T M$ to be precomputed in an offline step. Hence, with the values of the DEIM coordinates available, the ROM can be evaluated in a completely low-dimensional fashion.

Determining a function $c : \mathbb{R}^r \rightarrow \mathbb{R}^m$ to obtain the DEIM coordinates from the POD coordinates so that (20) holds exactly is an overdetermined problem. Furthermore, the least-squares solution to system (20), $c(a) = M^T C_h(\Phi a)$, requires an expensive projection operation. To solve this, the DEIM bases its approximation on only an m-dimensional subset of evaluations of $C_h(\Phi a)$ at points on the grid, referred to as the measurement points, and denoted by $\mathcal{P}$. More specifically, a measurement matrix $P \in \{0, 1\}^{N \times m}$ is defined that consists of selected columns of the $N \times N$ identity matrix corresponding to the measurement points. The DEIM coordinates are obtained from an interpolation condition of which the formal solution is $c(a) = (P^T M)^{-1} P^T C_h(\Phi a)$. A greedy measurement point selection procedure that ensures that $(P^T M)^{-1}$ exists and that provides near-optimal approximations of $C_h(\Phi a)$ was proposed in [24]. Using this method, the DEIM coordinates can be determined using only m evaluations of $C_h(\Phi a)$.

We will analyse the conservation properties of this 'conventional' DEIM approach. Because the telescoping property of operators discretized using the FVM is a linear property, it is conserved under application of the POD algorithm. This may be shown by analysing the eigenvectors of the correlation matrix $\Xi \Xi^T$. Using this property it can be shown that momentum is conserved for a hROM using the DEIM:

$$
\begin{aligned}
\left(\frac{d\boldsymbol{P}_r}{dt}\right)_i &= \frac{d}{dt}\left\langle \Phi^T \Omega_h e_i, a \right\rangle = -\left\langle \Phi^T \Omega_h e_i, \Phi^T M c(a) \right\rangle + \nu \left\langle \Phi^T \Omega_h e_i, D_r a \right\rangle \\
&= -\left\langle e_i, M c(a) \right\rangle + \nu \left\langle e_i, D_h \Phi a \right\rangle \\
&= 0.
\end{aligned}
\tag{21}
$$

We now arrive at the key issue addressed in this article. This issue is that *the hROM does not conserve reduced total kinetic energy*, i.e. it can occur that:

$$\frac{dK_r}{dt} = \frac{1}{2}\frac{d}{dt}\langle a, a \rangle = -\left\langle a, \Phi^T M c(a) \right\rangle - \nu \left\| Q_r a \right\|^2 > 0, \tag{22}$$

because generally $\left\langle a, \Phi^T M c(a) \right\rangle$ does not equal zero. This is because the DEIM-interpolated convection operator:

$$\Phi^T M (P^T M)^{-1} P^T \widetilde{C}_h(\Phi a)\Phi, \tag{23}$$

is no longer skew-symmetric. This can lead to instability as the norm of the generalized coordinates $a(t)$ is not bounded from above. An approach to retain the skew-symmetry of a DEIM approximation and therefore the energy-conservation property of the ROM has been proposed in [66]. However, this method scales equivalently to the exact tensor decomposition and is therefore considered infeasible for realistic fluid dynamics applications. We will present a new approach to make the hROM energy-conserving, while also retaining the momentum conservation property.

3.2. A novel energy- and momentum-conserving DEIM

The method we propose will be named the constrained least-squares discrete empirical interpolation method (CLSDEIM). It is based on the realization that skew-symmetry of the interpolated quasi-linear convection operator (23) is a sufficient *but not a necessary condition* for energy-conservation. Instead, a necessary condition to conserve reduced total kinetic energy is:

$$\left\langle a, \Phi^T M c(a) \right\rangle = 0. \tag{24}$$

Condition (24) can be satisfied even if the Galerkin-projected DEIM operator $\Phi^T M c : \mathbb{R}^r \rightarrow \mathbb{R}^r$ is not skew-symmetric and, when satisfied, results in the correct reduced kinetic energy evolution equation (17). The new idea of the CLSDEIM is to enforce this condition by posing the DEIM as a constrained minimization problem.

The conventional DEIM finds the DEIM coordinates $c(a)$ by minimizing the Euclidean norm between the nonlinearity and the DEIM approximation in the measurement space. The Euclidean norm can be considered as minimized since the difference in $\mathcal{P}$ i.e. $P^T C_h(\Phi a) - P^T M c(a)$, is zero. The idea of the proposed CLSDEIM is to employ this view of DEIM as a minimization problem and constrain it to take place over the set $\mathcal{F}(a)$ of DEIM approximations satisfying condition (24), defined using the DEIM coordinates as:

$$\mathcal{F}(a) := \{c \in \mathbb{R}^m \mid a^T \Phi^T M c = 0\}.$$

The set $\mathcal{F}(a)$ is referred to as the feasible set. As DEIM approximations with $c(a) \in \mathcal{F}(a)$ satisfy condition (24), the CLSDEIM produces approximations that conserve reduced total kinetic energy. The constrained minimization problem to find the DEIM coordinates $c(a)$ will be posed as the following linearly constrained least-squares problem:

$$c(a) = \underset{\widetilde{c} \in \mathbb{R}^m}{\arg\min} \left\| P^T C_h(\Phi a) - P^T M \widetilde{c} \right\|^2 \quad \text{s.t.} \quad a^T \Phi^T M \widetilde{c} = 0. \tag{25}$$

This means the CLSDEIM relaxes the interpolation condition in $\mathcal{P}$ between the FOM's convection operator and the DEIM approximation imposed by the conventional DEIM. Rather, the CLSDEIM minimizes the difference between the approximation and the FOM operator in $\mathcal{P}$ while simultaneously constraining the approximation to be energy-conserving.

The constrained minimization problem (25) is solved using the method of Lagrange multipliers [15]. The Lagrangian $\mathcal{L} : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ is defined as:

$$\mathcal{L}(c, \lambda) = \left\| P^T C_h(\Phi a) - P^T M c \right\|^2 + \lambda a^T \Phi^T M c,$$

where $\lambda \in \mathbb{R}$ is a Lagrange multiplier. Taking partial derivatives of the Lagrangian and setting them to zero leads to the following system for the optima $(c_0, \lambda_0) \in \mathbb{R}^m \times \mathbb{R}$:

$$\begin{bmatrix} 2(P^T M)^T P^T M & (a^T \Phi^T M)^T \\ a^T \Phi^T M & 0 \end{bmatrix} \begin{bmatrix} c_o \\ \lambda_o \end{bmatrix} = \begin{bmatrix} 2(P^T M)^T P^T C_h(\Phi a) \\ 0 \end{bmatrix}. \tag{26}$$

Note that this matrix is symmetric and has constant coefficients with the exception of the last row and column which depend on a. Solving (26) can be done explicitly using the inverse of a symmetric 2×2 block matrix [60]:

$$\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} & A^{-1} B (B^T A^{-1} B)^{-1} \\ (B^T A^{-1} B)^{-1} B^T A^{-1} & -(B^T A^{-1} B)^{-1} \end{bmatrix},$$

where A is assumed symmetric. Taking:

$$A := 2(P^T M)^T P^T M \qquad \in \mathbb{R}^{m \times m}$$

$$B := (a^T \Phi^T M)^T = M^T \Phi a \qquad \in \mathbb{R}^{m \times 1},$$

where A is indeed symmetric and nonsingular due to the construction of $\mathcal{P}$ [24], the optimal DEIM coordinates c_o solving (25) define the function c for the CLSDEIM coordinates as:

$$c(a) = \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) 2(P^T M)^T P^T C_h(\Phi a). \tag{27}$$

We use the notation $B := b(a)$ to explicitly denote the dependence of B on a and the fact that it can be treated as a vector. Since $B^T A^{-1} B = b(a)^T A^{-1} b(a)$ and $B^T A^{-1} 2(P^T M)^T P^T C_h(\Phi a) = b(a)^T (P^T M)^{-1} P^T C_h(\Phi a)$ are simply scalars, the final form of the function c is:

$$\text{CLSDEIM:} \qquad c(a) = (P^T M)^{-1} P^T C_h(\Phi a) - \frac{b(a)^T (P^T M)^{-1} P^T C_h(\Phi a)}{b(a)^T A^{-1} b(a)} A^{-1} b(a). \tag{28}$$

The DEIM coordinates found from equation (28) result in a DEIM approximation $M c(a)$ satisfying condition (24) and consequently also equation (17). All terms in (28) are either inner products or matrix-vector products between low-dimensional vectors or matrices and vectors respectively, hence the CLSDEIM can be solved in a low-dimensional fashion. Assuming that $m \geq r$ and that $(P^T M)^{-1}$ and A^{-1} are precomputed, the most expensive operations in (28) are calculating $(P^T M)^{-1} P^T C_h(\Phi a)$ once $P^T C_h(\Phi a)$ is obtained and $A^{-1} b(a)$ once $b(a)$ is obtained. The computational complexity of both steps scales identically as $\mathcal{O}(m^2)$. Evaluating the DEIM coordinates using the CLSDEIM therefore scales computationally as $\mathcal{O}(m^2)$, the same scaling as the conventional DEIM. Note that the first term on the right hand side of equation (28) is the solution to the DEIM interpolation condition. Hence, equation (28) can be interpreted as the conventional DEIM with an extra term arising from the constraint. The extra term projects the conventional DEIM coordinates on $\ker(\langle b(a), \cdot \rangle)$ such that condition (24) is satisfied and minimization problem (25) is solved.

Considering the objective function and the geometric interpretation of the DEIM [26,24], the CLSDEIM can be interpreted geometrically as an oblique projection of $C_h(\Phi a)$ on the subspace $\mathcal{M}_{\mathcal{F}} \subset \mathbb{R}^N$ of all DEIM approximations $M \widetilde{c}$ with $\widetilde{c} \in \mathcal{F}(a)$. The subspace $\mathcal{M}_{\mathcal{F}}$ is defined as:

$$\mathcal{M}_{\mathcal{F}} = \mathcal{M} \cap \ker(\langle \Phi a, \cdot \rangle),$$

where $\langle \Phi a, \cdot \rangle : \mathbb{R}^N \to \mathbb{R}$ is the functional taking the Euclidean inner product with Φa. Particularly, $\mathcal{M}_{\mathcal{F}}$ are all vectors in the range of M that satisfy condition (24). Since $\mathcal{M}_{\mathcal{F}}$ is the intersection between two linear subspaces of $\mathbb{R}^N$ it is also a linear subspace of $\mathbb{R}^N$. Contrary to the conventional DEIM, the CLSDEIM also projects obliquely through the measurement space $\mathcal{P}$. Hence, the orthogonal projection of the DEIM residual $r(t) \in \mathbb{R}^N$ on $\mathcal{P}$, as given by:

$$P^T r(t) = P^T \left[C_h(\Phi a(t)) - M c(a(t)) \right],$$

is generally not zero. However, the CLSDEIM projector $\Pi_{\mathcal{M}_{\mathcal{F}}} : \mathbb{R}^N \to \mathcal{M}_{\mathcal{F}}$ does project $C_h(\Phi a)$ onto $\mathcal{M}_{\mathcal{F}}$ such that the orthogonally projected residual $P^T r(t)$ is minimal (with respect to the elements of $\mathcal{F}(a)$) in the Euclidean norm. This statement follows trivially

from the formulation of the CLSDEIM minimization problem (25). The geometrical interpretation of this statement is that, in the space $\mathbb{R}^m$, the CLSDEIM calculates the coefficients with respect to the (non-orthogonal) basis $P^T M$ of the orthogonal projection of $P^T C_h(\Phi a)$ onto the subspace $\ker\left(\left\langle M^T \Phi a, (P^T M)^{-1}(\cdot)\right\rangle\right) \subset \mathbb{R}^m$. Finally, a proof that $\Pi_{\mathcal{M}_r}$ is in fact an oblique projection operator, existence and uniqueness results for CLSDEIM minimizers and a proof of some error bounds are provided in Appendix B, Appendix A and Appendix C, respectively. The error bounds show that the CLSDEIM is also nearly optimal in the same sense as described for the conventional DEIM in [24].

Furthermore, the CLSDEIM basis M and the measurement space $\mathcal{P}$ can simply be found following the procedures of the conventional DEIM algorithm. Indeed, as we are using the conventional DEIM basis the reduced total momentum will also remain a conserved quantity for the CLSDEIM as was shown in equation (21).

Remark 1. Both implicit and explicit Runge-Kutta (RK) methods will be considered to integrate the proposed hROMs. To achieve exact energy conservation (for inviscid flows) energy-conserving RK methods can be used. Many families of energy-conserving RK methods are *implicit* [86,87], however *explicit* energy-conserving RK methods exist too (see [51,79] and the references therein). Using such methods and energy-conserving hyper-reduction, the change in reduced total kinetic energy K_r is given by:

$$K_r^{n+1} - K_r^n = -\Delta t \sum_{i=1}^s \beta_i \nu \left|\left| Q_r A_i \right|\right|^2, \tag{29}$$

which is the time-discrete equivalent of (17). A derivation of (29) has been added in Appendix D. Here, $s \in \mathbb{N}$ and $\beta_i \in \mathbb{R}$ are parameters of the given RK method and $A_i \in \mathbb{R}^r$ are the generalized coordinates at the ith RK stage. In the inviscid case it can be seen from (29) that the hROMs conserve reduced kinetic energy. In the viscous case the resulting hROMs will be nonlinearly stable when $\beta_i \geq 0 \ \forall \ i \in \{1,...,s\}$, which holds for example for the Gauss-Legendre RK methods (see [87,86] for further details).

3.3. Preventing overfitting errors

Interpolation-based methods like the DEIM and CLSDEIM can be prone to overfitting when noise is present [44]. It was shown in [75] that adding Gaussian noise $\epsilon \in \mathbb{R}^m$ to the measurements, i.e. $P^T C_h(u) + \epsilon$, results in error bounds on the DEIM approximation of $C_h(u)$ that grow as $\mathcal{O}(\sqrt{m})$. Similar overfitting errors may occur in the presence of numerical errors in a as we will show using experiments later. As a result of such overfitting errors, numerical errors in the solution may accumulate, further harming the capability of the DEIM to reconstruct the correct $C_h(\Phi a)$ of the following time steps. This is especially notable when the interpolation procedure is carried out many times, which is the case when solving time-dependent problems. We remark that this can happen even in the case of fully-discrete energy-conserving schemes since energy conservation is a restriction on the norm of the solution and not on its accuracy. In this case, overfitting errors typically manifest themselves in non-convergence of the nonlinear solver. To address overfitting we therefore propose two further stabilization approaches. The first one is oversampling, which has recently gained significant attention [75,83]. The second approach is regularization, which is a well-known technique from optimization and has thus far, to the author's knowledge, not been applied to the DEIM. This latter approach has the benefit that additional measurement points are not necessary and therefore reduces the computational burden compared to oversampling.

3.3.1. Oversampling

When using oversampling we replace the interpolation of the DEIM and CLSDEIM (25) with a regression procedure. This is achieved by considering a larger measurement space than the dimension of the DEIM space $\mathcal{M}$, thus $m_p := \dim(\mathcal{P}) > m$. As a result, the measurement matrix becomes $P \in \{0,1\}^{N \times m_p}$. The formulation of the underlying optimization problem remains the same as optimization problem (25), but now the larger measurement matrix is used. Due to the fact that we are trying to fit a linear combination of m DEIM modes to m_p measurements with $m_p > m$, the resulting DEIM approximation $Mc(a)$ is less sensitive to noise in measurements of $C_h(\Phi a)$ in $\mathcal{P}$. This was also analysed in [75], where it was shown that oversampling using randomized measurement points resulted in a contribution of noise to the error bound of the approximation of $C_h(u)$ that scales with $\mathcal{O}\left(\sqrt{m/m_p}\right)$. Hence, overfitting errors as a result of noise can be mitigated by choosing m_p sufficiently large compared to m. We will refer to CLSDEIM being extended with oversampling as OCLSDEIM. The DEIM coordinates as calculated using the OCLSDEIM require solving optimization problem (25) with the new measurement matrix P. The solution procedure of this problem is completely analogous to that of the CLSDEIM (25) and the final expression for the optimal c_o defines the OCLSDEIM function:

$$\text{OCLSDEIM:} \quad c(a) = (P^T M)^\dagger P^T C_h(\Phi a) - \frac{b(a)^T (P^T M)^\dagger P^T C_h(\Phi a)}{b(a)^T A^{-1} b(a)} A^{-1} b(a), \tag{30}$$

where instead of $(P^T M)^{-1}$ the Moore-Penrose pseudoinverse $(P^T M)^\dagger$ appears. Assuming $m \geq r$ and that $(P^T M)^\dagger \in \mathbb{R}^{m \times m_p}$ and A^{-1} are precomputed, the most expensive step in evaluating (30) is computing $(P^T M)^\dagger P^T C_h(\Phi a)$ once $P^T C_h(\Phi a)$ is obtained. As a result, calculating the DEIM coordinates using the OCLSDEIM scales as $\mathcal{O}(m_p m)$.

The measurement point selection procedure will be as follows. The first m measurement points are found using the classic DEIM procedure described in [24]. The next $m_p - m$ points are generated from a uniform random distribution where care is taken to

not select repeating indices. This procedure strikes a good balance between efficiency and accuracy and is simple to implement. Alternatives are possible, see for example the following articles for deterministic approaches to measurement point selection: [75,83, 100,61].

3.3.2. Generalized Tikhonov regularization

When overfitting occurs, DEIM modes interpolate measurements containing noise or numerical errors resulting from time integration, negatively effecting accuracy in subsequent time steps. Although the DEIM measurement points are chosen to minimize $||(P^T M)^{-1}||$ thereby mitigating the effect of overfitting [24], accumulation of errors can still take place as we will see during the numerical experiments. This accumulation can eventually result in severe errors and even numerical instability. To prevent this issue we propose regularizing the least-squares problem using a generalized Tikhonov regularization [44]. Using such a regularization procedure we can penalize undesired properties of solutions to the DEIM optimization problem by adding extra terms to its associated cost function. In this article we will propose to use regularization terms of the form:

$$c(a) = \underset{\widetilde{c} \in \mathbb{R}^m}{\arg\min} \left|\left| P^T C_h(\Phi a) - P^T M \widetilde{c} \right|\right|^2 + \alpha \left|\left| \widetilde{c} - \mu \right|\right|_\Gamma^2 \quad \text{s.t.} \quad a^T \Phi^T M \widetilde{c} = 0. \tag{31}$$

Here $||\tilde{c}||_\Gamma := \sqrt{\langle \tilde{c}, \Gamma \tilde{c} \rangle}$ with $\Gamma \in \mathbb{R}^{m \times m}$ symmetric positive definite (SPD), $\mu \in \mathbb{R}^m$ is some desired or reference state and $\alpha \in \mathbb{R}^+$ is a hyperparameter. The minimization problem can again be solved using the method of Lagrange multipliers. The Lagrangian $\mathcal{L} : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ of this minimization is defined as:

$$\mathcal{L}(c, \lambda; \alpha) = \left|\left| P^T C_h(\Phi a) - P^T M c \right|\right|^2 + \alpha ||c - \mu||_\Gamma^2 + \lambda a^T \Phi^T M c,$$

where $\lambda \in \mathbb{R}$ is a Lagrange multiplier. Taking partial derivatives of the Lagrangian and setting them to zero leads to the following system for the optima $(c_0, \lambda_0) \in \mathbb{R}^m \times \mathbb{R}$:

$$\begin{bmatrix} 2\left((P^T M)^T P^T M + \alpha\Gamma\right) & (a^T \Phi^T M)^T \\ a^T \Phi^T M & 0 \end{bmatrix} \begin{bmatrix} c_o \\ \lambda_o \end{bmatrix} = \begin{bmatrix} 2(P^T M)^T P^T C_h(\Phi a) + 2\alpha\Gamma\mu \\ 0 \end{bmatrix}.$$

Again, this matrix is symmetric and has constant coefficients with the exception of the last row and column which depend on a. Denoting:

$$A_\Gamma := (P^T M)^T P^T M + \alpha\Gamma \in \mathbb{R}^{m \times m},$$

the DEIM coordinates as calculated by the regularized CLSDEIM are given as:

Regularized CLSDEIM:

$$c(a) = A_\Gamma^{-1}\left((P^T M)^T P^T C_h(\Phi a) + \alpha\Gamma\mu\right) - \frac{b(a)^T A_\Gamma^{-1}\left((P^T M)^T P^T C_h(\Phi a) + \alpha\Gamma\mu\right)}{b(a)^T A_\Gamma^{-1} b(a)} A_\Gamma^{-1} b(a). \tag{32}$$

Note that this equation also applies to a combination of oversampling and regularization. Taking $\alpha = 0$ in equation (32) we obtain the normal CLSDEIM since $A_\Gamma^{-1}(P^T M)^T = (P^T M)^\dagger = (P^T M)^{-1}$ in this case. A data-driven method to calculate Γ and μ that uses the knowledge we have on the DEIM coordinates of our snapshot data Ξ will be proposed in the following section. The hyperparameter tuning of α was carried out manually for simplicity. More advanced approaches to find a priori well-performing hyperparameters for regularized minimization problems like (31) are generalized cross-validation [35] and L-curve based methods [42], which we suggest implementing in future work. Assuming $m \geq r$, that A_Γ^{-1} and $P^T M$ are precomputed and that we are combining oversampling and Tikhonov regularization, the most expensive step in evaluating (32) is calculating $(P^T M)^T P^T C_h(\Phi a)$ once $P^T C_h(\Phi a)$ is obtained. Due to oversampling the computational complexity of the previous step can generally be $\mathcal{O}(m_p m)$, where $m_p = m$ when we are strictly using Tikhonov regularization.

3.3.3. Mahalanobis distance

The method we propose to determine the regularization term is to use the Mahalanobis distance $d_M : \mathbb{R}^m \to \mathbb{R}^+$ [44] of the solution $c(a)$ with respect to the data matrix $M^T \Xi \in \mathbb{R}^{m \times n_s}$. The Mahalanobis distance is computed as:

$$d_M(c) = \left|\left| c - \mu_c \right|\right|_{S^{-1}},$$

and thus corresponds to taking $\Gamma = S^{-1}$ and $\mu = \mu_c$ in optimization problem (31). Here, the matrix $S \in \mathbb{R}^{m \times m}$ and vector $\mu_c \in \mathbb{R}^m$ are the covariance matrix and the mean associated to the data in $M^T \Xi$, respectively. This method is based on the idea of promoting solutions $c(a)$ that are similar to those observed in the data matrix $M^T \Xi$ (which has the DEIM coordinates of the snapshots in Ξ as its columns). Here, similar is in the sense of originating from the same distribution. In fact, regularization using the Mahalanobis distance d_M corresponds to assuming a prior distribution on the optimal $c(a)$ that can be accurately characterized by a mean μ_c and a covariance matrix S. The Mahalanobis distance is then the multidimensional generalization of the number of standard deviations an observation is removed from the distribution's mean.

In this paper we will concern ourselves with solution reproduction problems for simplicity, and not with parameterized problems. We will see that, as numerical errors accumulate during time integration, DEIM coordinates take increasingly different values from their reference values in $M^T \Xi$ when using the DEIM or CLSDEIM. Hence, such values will contribute to a large $d_M(c(a))$ and as a result less optimal than a more well-behaved solution given the regularized CLSDEIM loss-function (31). We assume that this idea can be extended to the parametric case, i.e. that the DEIM coordinates of a parametric problem can also be treated as generated by a single prior distribution. We will refer to the CLSDEIM regularized by the Mahalanobis distance as the MCLSDEIM.

The Mahalanobis distance $d_M(c(a))$ and the residual term in (31) can differ by orders of magnitude. In the current work we solve this issue by taking very small hyperparameter values α. We will validate the choice for such a small hyperparameter for one of our experiments in the results section by means of plotting a number of L-curves [42] and showing that the chosen hyperparameter produces solutions in the heuristically optimal L-curve corner.

Remark 2. L-curves provide a heuristic to choose good hyperparameters for (generalized) Tikhonov regularization terms in least-squares problems [42]. They are curves parameterized by the hyperparameter $\alpha \in \mathbb{R}^+$ tracing the quantities $(||P^T C_h(\Phi a) - P^T M c_\alpha||, ||c_\alpha - \mu||_\Gamma)$ as in (31) with $c_\alpha \in \mathbb{R}^m$ being the solution to (31) for a specific hyperparameter value α. L-curves are typically given in log-log plots. When α becomes large L-curves for problems like (31) tend to level off as the regularization term $||c_\alpha - \mu||_\Gamma$ begins to dominate the solution and the residual term $||P^T C_h(\Phi a) - P^T M c_\alpha||$ increases due to a bad fit to the data. This results in a horizontal region in the L-curve. When α becomes small the residual term decreases, but the regularization term increases sharply due to bad properties of the fit, resulting in a vertical region. The vertical and horizontal parts of these L-curves are connected by a corner where the heuristic indicates that the residual term and regularization term are balanced. Intuitively, this makes the range of associated hyperparameters in the L-curve corner good choices. We suggest [42,41,56] for a more detailed discussion on L-curves and Tikhonov regularization.

4. An energy- and momentum-conserving temporal localization method

Having constructed new energy-conserving and robust DEIM methods, we now extend the feasibility of our approach with a new energy- and momentum-conserving temporal localization method. Namely, it is well known that as flows become more convection dominated ($\nu \to 0$) the Kolmogorov N-width decay becomes increasingly slower. This makes it more difficult for conventional POD to generate a low-dimensional basis set that accurately captures a significant amount of energy contained in both Ξ and the solution snapshots $X \in \mathbb{R}^{N \times n_s}$. To deal with this problem researchers have proposed several solutions, one of which is the application of the principal interval decomposition (PID) for both the construction of DEIM and POD spaces, $\mathcal{M}$ and $\mathcal{V}$ respectively [22,74,49,14,3,38, 27]. However, as far as we know energy and momentum conservation has not been taken into account in existing PID approaches. In what follows, we will propose a method to preserve energy and momentum throughout a full time integration using a hROM with temporal localization.

4.1. Temporal localization using PID

The PID decomposes the snapshot sets over n intervals in time $[t_i, t_{i+1}]$ and applies the POD algorithm to the individual intervals. The idea is that by calculating modes tailored to specific intervals, the local timescales within the respective intervals are captured significantly better than by a set of modes calculated from the full set of snapshots [14,22]. Based on snapshot sets:

$$X = [X_1, X_2, ..., X_n], \quad \Xi = [\Xi_1, \Xi_2, ..., \Xi_n],$$

the PID provides sets of POD modes:

$$\Phi_i, M_i, \ i \in \{1, ..., n\}, \tag{33}$$

applicable to use at times $t \in [t_i, t_{i+1}]$ within their respective intervals. The local POD modes Φ_i follow from an SVD of the local snapshot matrix $X_i \in \mathbb{R}^{N \times n_s^i}$, where $n_s^i \in \mathbb{N}$ is the number of snapshots in the ith interval. The temporally localized DEIM measurement space $\mathcal{P}_i$ for the ith interval is determined solely based on the DEIM modes in M_i using the algorithm as in [24] specifically for M_i. Furthermore, we have $\mathcal{V}^i := \text{span}(\Phi_i) \subset \mathbb{R}^N$ and $\mathcal{M}^i := \text{span}(M_i) \subset \mathbb{R}^N$.

Setting up the hROM using the PID, the dynamical system takes the form:

$$\frac{d a_i}{dt} = -\Phi_i^T M_i c_i(a_i) + \nu D_r^i a_i \quad t \in [t_i, t_{i+1}], \tag{34}$$

where $D_r^i := \Phi_i^T D_h \Phi_i$. Note that we also introduce subscripts for the generalized and DEIM coordinates a_i and c_i respectively, as they are only valid during the interval $[t_i, t_{i+1}]$. The DEIM coordinates are calculated using any of the previously discussed methods using the appropriate measurement spaces and DEIM and POD bases. It should be noted that all the conservation properties of the energy- and momentum-conserving DEIM methods hold *within* intervals. This is trivial as nothing changes in the way the DEIM coordinates are calculated. Instead, the art is to develop interface conditions such that during the transition from one interval to the next, these conservation properties are still satisfied.

Note that in the subsequent sections a hROM using the PID and a certain DEIM algorithm will be referred to as PID-CLSDEIM hROM (when for example the CLSDEIM is used without regularization).

4.2. Interface conditions

The bases in (33) are used only during the interval $t \in [t_i, t_{i+1}]$. At $t = t_{i+1}$ an interface is reached and it is required to transition from a solution described in terms of $\boldsymbol{u}_r^- := \Phi_i \boldsymbol{a}_i(t_{i+1}) \in \mathcal{V}^i$ to one described in terms of $\boldsymbol{u}_r^+ := \Phi_{i+1} \boldsymbol{a}_{i+1}(t_{i+1}) \in \mathcal{V}^{i+1}$, where $\boldsymbol{a}_i(t) \in \mathbb{R}^{r_i}$, $\boldsymbol{a}_{i+1}(t) \in \mathbb{R}^{r_{i+1}}$ and in general $\dim(\mathcal{V}^i) = r_i \neq r_{i+1} = \dim(\mathcal{V}^{i+1})$. During the numerical experiments in this article we will consider the case $r_i = r \; \forall \, i \in \{1, ..., n\}$. To determine how to divide the solution manifold approximation over different linear subspaces, one could use heuristic methods [22] or a clustering algorithm [38], but this is outside the scope of this article.

We will define how to transition between intervals based on an interface condition. Using the conventional PID [3] the following interface condition is defined:

$$\text{find} \quad \boldsymbol{a}_{i+1}(t_{i+1}) \quad \text{s.t.} \quad \left\langle \boldsymbol{u}_r^+ - \boldsymbol{u}_r^-, \boldsymbol{v} \right\rangle_{\Omega_h} = 0 \quad \forall \, \boldsymbol{v} \in \mathcal{V}^{i+1}, \tag{35}$$

which is stable in the sense that it does not lead to an energy increase, as we will show. This condition is satisfied by the coefficients with respect to Φ_{i+1} of the Ω_h-orthogonal projection of $\boldsymbol{u}_r^-$ onto $\mathcal{V}^{i+1}$:

$$\boldsymbol{a}_{i+1}(t_{i+1}) = \Phi_{i+1}^T \Omega_h \boldsymbol{u}_r^- = \Phi_{i+1}^T \Omega_h \Phi_i \boldsymbol{a}_i(t_{i+1}). \tag{36}$$

Note that the interface matrix $T_i := \Phi_{i+1}^T \Omega_h \Phi_i \in \mathbb{R}^{r_{i+1} \times r_i}$ can be precomputed in an offline stage to obtain a cheap online evaluation. The kinetic energy $K_r^+ := \frac{1}{2} ||\boldsymbol{u}_r^+||_{\Omega_h}^2 = \sum_k \frac{1}{2} \left\langle \boldsymbol{u}_r^-, (\Phi_{i+1})_{,k} \right\rangle_{\Omega_h}^2$ cannot increase beyond $K_r^- := \frac{1}{2} ||\boldsymbol{u}_r^-||_{\Omega_h}^2$ due to the Bessel inequality [54], which states that for the (complete) inner product space defined as $\mathcal{H}_{\Omega_h} := \left(\mathbb{R}^N, \langle \cdot, \cdot \rangle_{\Omega_h} \right)$ and orthonormal sequence $\left((\Phi_{i+1})_{,k} \right)$ with $(\Phi_{i+1})_{,k} \in \mathcal{H}_{\Omega_h}$, $k \in \{1, ..., r_{i+1}\}$ we have the following:

$$\text{Bessel Inequality in } \mathcal{H}_{\Omega_h}: \quad \sum_{k=1}^{r_{i+1}} \left\langle \boldsymbol{u}_r^-, (\Phi_{i+1})_{,k} \right\rangle_{\Omega_h}^2 \leq ||\boldsymbol{u}_r^-||_{\Omega_h}^2 .$$

Hence, application of the conventional PID using interface condition (35) cannot lead to an increase in kinetic energy for any number of intervals n, making it *nonlinearly stable*. Furthermore, the solution is also optimal in the sense that (36) is also the solution to the minimization problem:

$$\boldsymbol{a}_{i+1}(t_{i+1}) = \underset{\boldsymbol{a} \in \mathbb{R}^{r_{i+1}}}{\arg\min} ||\boldsymbol{u}_r^- - \Phi_{i+1} \boldsymbol{a}||_{\Omega_h}^2 . \tag{37}$$

In summary, the conventional PID is nonlinearly stable. However, kinetic energy and momentum are in general not exactly conserved when transitioning between intervals using condition (35), which can lead to artificial (numerical) dissipation of the solution. To solve this issue of lack of conservation, we propose the following new energy- and momentum-conserving interface condition.

4.3. Energy- and momentum-conserving interface conditions

To consider energy and momentum conservation over interfaces we define $\left(\boldsymbol{P}_r^+ \right)_k := \left\langle \boldsymbol{e}_k, \boldsymbol{u}_r^+ \right\rangle_{\Omega_h}$ and similarly $\left(\boldsymbol{P}_r^- \right)_k := \left\langle \boldsymbol{e}_k, \boldsymbol{u}_r^- \right\rangle_{\Omega_h}$. In order to conserve energy and momentum over interfaces we benefit from the optimization formulation (37) underlying the non-conserving interface condition (35). We propose to constrain the optimization problem (37) to conserve $K_r(t)$ and $\boldsymbol{P}_r(t)$, leading to:

$$\boldsymbol{a}_{i+1}(t_{i+1}) = \underset{\boldsymbol{a} \in \mathbb{R}^{r_{i+1}}}{\arg\min} ||\boldsymbol{u}_r^- - \Phi_{i+1} \boldsymbol{a}||_{\Omega_h}^2 \quad \text{s.t.} \quad \frac{1}{2} ||\boldsymbol{a}||^2 = K_r^-, \quad \left\langle \boldsymbol{e}_k, \Phi_{i+1} \boldsymbol{a} \right\rangle_{\Omega_h} = \left(\boldsymbol{P}_r^- \right)_k \forall k \in \{1, ..., d\}. \tag{38}$$

The constraints on this problem can be simplified by realizing that the momentum component $(\boldsymbol{P}_r(t))_k$ in direction k is fully determined by the kth component $(\boldsymbol{a}(t))_k$ of the generalized coordinates, namely:

$$\left(\boldsymbol{P}_r(t) \right)_k = \left\langle \boldsymbol{e}_k, \sum_{j=1}^r \Phi_{,j} (\boldsymbol{a}(t))_j \right\rangle_{\Omega_h} = \left\langle \boldsymbol{e}_k, \frac{\boldsymbol{e}_k}{||\boldsymbol{e}_k||_{\Omega_h}} (\boldsymbol{a}(t))_k \right\rangle_{\Omega_h} = ||\boldsymbol{e}_k||_{\Omega_h} (\boldsymbol{a}(t))_k,$$

due to Ω_h-orthogonality. Furthermore, the first d POD modes for every interval are the same and given by the momentum-conserving modes $\boldsymbol{e}_k / ||\boldsymbol{e}_k||_{\Omega_h}$. As a result, the momentum-conserving interface constraint can only be satisfied by setting $(\boldsymbol{a}_{i+1}(t_{i+1}))_k = (\boldsymbol{a}_i(t_{i+1}))_k \; \forall k \in \{1, ..., d\}$. The following matrices and vectors are now defined:

$$\widetilde{\Phi}_i := \left[(\Phi_i)_{,d+1}, \; ... \; , (\Phi_i)_{,r_i} \right], \quad \widetilde{\boldsymbol{a}}_i(t) := \left[(\boldsymbol{a}_i(t))_{d+1}, \; ... \; , (\boldsymbol{a}_i(t))_{r_i} \right]^T .$$

Since the momentum conservation constraint fully determines the first d components of the generalized coordinates $\boldsymbol{a}_{i+1}(t)$ we can simplify the optimization problem (38) by formulating it in terms of the new $\widetilde{(\cdot)}$ variables. Note that at any interface:
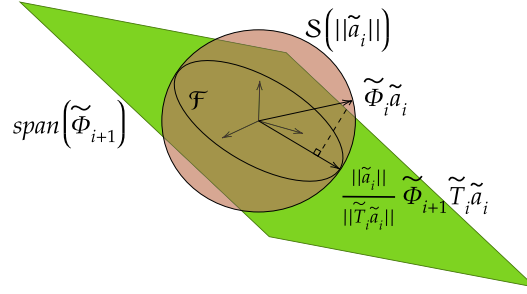
Fig. 1. Graphical intuition for optimization solution with $N = 3$ and $\Omega_h = I$.

$$
\boldsymbol{u}_r^+ - \boldsymbol{u}_r^- = \sum_{k=d+1}^{r_{i+1}} (\Phi_{i+1})_{,k} (\boldsymbol{a}_{i+1}(t_{i+1}))_k - \sum_{l=d+1}^{r_i} (\Phi_i)_{,l} (\boldsymbol{a}_i(t_{i+1}))_l + \sum_{j=1}^{d} \frac{\boldsymbol{e}_j}{||\boldsymbol{e}_j||_{\Omega_h}} \left((\boldsymbol{a}_{i+1}(t_{i+1}))_j - (\boldsymbol{a}_i(t_{i+1}))_j \right)
$$

$$
= \widetilde{\Phi}_{i+1} \widetilde{\boldsymbol{a}}_{i+1}(t_{i+1}) - \widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})
$$

due to the momentum-conserving interface conditions. The kinetic energy constraint can also be simplified to $||\widetilde{\boldsymbol{a}}_{i+1}(t_{i+1})||^2 = ||\widetilde{\boldsymbol{a}}_i(t_{i+1})||^2$ since the first d components of the generalized coordinates $\boldsymbol{a}_{i+1}(t_{i+1})$ and $\boldsymbol{a}_i(t_{i+1})$ are the same. The reduced minimization problem for the remaining $r_{i+1} - d$ components of the generalized coordinates $\boldsymbol{a}_{i+1}(t_{i+1})$ then becomes:

$$
\widetilde{\boldsymbol{a}}_{i+1}(t_{i+1}) = \underset{\boldsymbol{a} \in \mathbb{R}^{r_{i+1}-d}}{\arg\min} \left\| \widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) - \widetilde{\Phi}_{i+1} \boldsymbol{a} \right\|_{\Omega_h}^2 \quad \text{s.t.} \quad ||\boldsymbol{a}||^2 = ||\widetilde{\boldsymbol{a}}_i(t_{i+1})||^2 . \tag{39}
$$

As before, we solve this constrained minimization problem using the method of Lagrange multipliers [15]. The Lagrangian $\mathcal{L}$: $\mathbb{R}^{r_{i+1}-d} \times \mathbb{R} \to \mathbb{R}$ of this minimization is as follows:

$$
\mathcal{L}(\boldsymbol{a}, \lambda) = \boldsymbol{a}^T \boldsymbol{a} - 2\boldsymbol{a}^T \widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) + \lambda \left(\boldsymbol{a}^T \boldsymbol{a} - \widetilde{\boldsymbol{a}}_i(t_{i+1})^T \widetilde{\boldsymbol{a}}_i(t_{i+1}) \right),
$$

where $\widetilde{T}_i := \widetilde{\Phi}_{i+1}^T \Omega_h \widetilde{\Phi}_i$ and $\lambda \in \mathbb{R}$ is a Lagrange multiplier. Taking partial derivatives of the Lagrangian and setting them to zero leads to the following nonlinear system for the optima $(\boldsymbol{a}_o, \lambda_o)$:

$$
\nabla_{\boldsymbol{a}} \mathcal{L} = (1 + \lambda_o) \boldsymbol{a}_o - \widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) = 0
$$

$$
\nabla_\lambda \mathcal{L} = \boldsymbol{a}_o^T \boldsymbol{a}_o - \widetilde{\boldsymbol{a}}_i(t_{i+1})^T \widetilde{\boldsymbol{a}}_i(t_{i+1}) = 0.
$$

Rewriting the equation in the first line, an expression for the generalized coordinates at the optimum $\boldsymbol{a}_o$ is obtained. This expression can be substituted into the second optimality condition to solve for $\frac{1}{1+\lambda_o}$. Finally, this results in:

$$
\widetilde{\boldsymbol{a}}_{i+1}(t_{i+1}) = \boldsymbol{a}_o = \frac{||\widetilde{\boldsymbol{a}}_i(t_{i+1})||}{\left\| \widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) \right\|} \widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}). \tag{40}
$$

To interpret this solution, we consider the subset $S(\delta) := \{ \boldsymbol{u} \in \mathcal{H}_{\Omega_h} \mid ||\boldsymbol{u}||_{\Omega_h} = \delta \}$ of all vectors in $\mathcal{H}_{\Omega_h}$ with the same norm of $\delta \in \mathbb{R}^+$ (and thus the same kinetic energy) and define the feasible set of minimization problem (39):

$$
\mathcal{F}_a := \{ \boldsymbol{a} \in \mathbb{R}^{r_{i+1}-d} \mid \widetilde{\Phi}_{i+1} \boldsymbol{a} \in S(||\widetilde{\boldsymbol{a}}_i(t_{i+1})||) \}
$$

and the set $\mathcal{F} := \text{span}(\widetilde{\Phi}_{i+1}) \cap S(||\widetilde{\boldsymbol{a}}_i(t_{i+1})||) = \{ \widetilde{\Phi}_{i+1} \boldsymbol{a} \mid \boldsymbol{a} \in \mathcal{F}_a \}$. The solution then corresponds to an orthogonal projection (in $\mathcal{H}_{\Omega_h}$) of $\widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})$ on $\text{span}(\widetilde{\Phi}_{i+1})$ and a subsequent rescaling to obtain an element of $\mathcal{F}$. Solution (40) then produces the generalized coordinates with respect to $\widetilde{\Phi}_{i+1}$ of this scaled projection. A visual interpretation for $N = 3$ and $\Omega_h = I$ is provided in Fig. 1.

Using equation (40) and the momentum conservation constraint the generalized coordinates of $\boldsymbol{u}_r^+$ as determined by the fully constrained minimization problem (38) are given by:

$$
\boldsymbol{a}_{i+1}(t_{i+1}) = \left[(\boldsymbol{a}_i(t_{i+1}))_1 \,,\, \dots \,,\, (\boldsymbol{a}_i(t_{i+1}))_d \,,\, \left(\frac{||\widetilde{\boldsymbol{a}}_i(t_{i+1})||}{||\widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})||} \widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) \right)^T \right]^T .
$$

We will refer to using the PID with this energy- and momentum-conserving interface condition as SP-PID.

It may happen that $||\widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})|| = 0$, in this case we cannot scale the result to have the same norm, and thus kinetic energy, as $\widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})$. This situation occurs when $\widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) \perp_{\Omega_h} \text{span}(\widetilde{\Phi}_{i+1})$ and it can be shown that, if $\dim(\text{span}(\widetilde{\Phi}_{i+1})), N > 1$, minimization problem (39) has an infinite number of solutions. Namely, writing out the loss function of minimization problem (39), taking into account that $\widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) \perp_{\Omega_h} \text{span}(\widetilde{\Phi}_{i+1})$), we have $||\widetilde{\Phi}_i \widetilde{\boldsymbol{a}}_i(t_{i+1}) - \widetilde{\Phi}_{i+1} \boldsymbol{a}||_{\Omega_h} = \sqrt{2} ||\widetilde{\boldsymbol{a}}_i(t_{i+1})|| \; \forall \; \boldsymbol{a} \in \mathcal{F}_a$ and since $|\mathcal{F}_a|$ is infinite in this
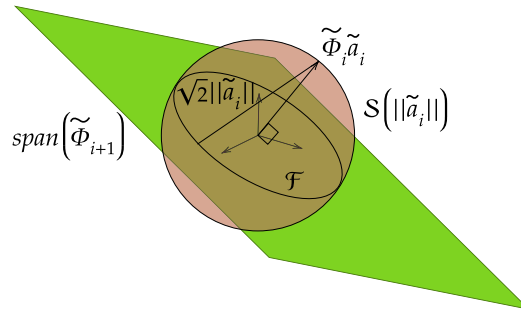
Fig. 2. Graphical intuition for infinite number of optimization solutions with $N = 3$ and $\Omega_h = I$.
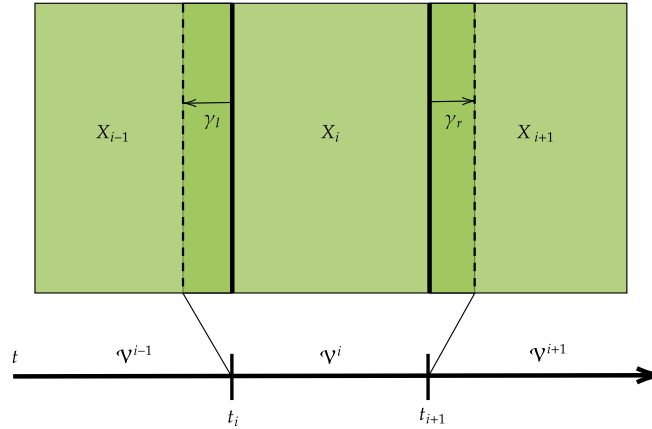

Fig. 3. Constructing $\mathcal{V}^i$ using overlapping snapshots.

case we have an infinite number of solutions. A visual representation of this case for $N = 3$ and $\Omega_h = I$ is given in Fig. 2. In principle any element of $\mathcal{F}_a$ can be used to determine $\boldsymbol{u}_r^+$ in this situation. We note that in general $||\widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})|| = 0$ is an unlikely scenario since the snapshots at the end of X_i tend to resemble those at the beginning of X_{i+1}. As a result, $\mathcal{V}^i$ and $\mathcal{V}^{i+1}$ are typically capable of resolving snapshots close to either side of their shared interface. However, to make sure that we prevent situations where $||\widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})|| = 0$ (or close to zero) we suggest to use overlapping snapshots. This entails concatenating $\gamma_l \in \mathbb{N}$ of the rightmost snapshots in X_{i-1}, Ξ_{i-1} and $\gamma_r \in \mathbb{N}$ of the leftmost snapshots in X_{i+1}, Ξ_{i+1} with X_i and Ξ_i respectively for the construction of Φ_i and M_i, as illustrated in Fig. 3. This will promote non-empty and larger intersections between the reduced spaces and as a result discourage $||\widetilde{T}_i \widetilde{\boldsymbol{a}}_i(t_{i+1})|| = 0$, because both reduced spaces $\mathcal{V}^i$ and $\mathcal{V}^{i+1}$ are constructed to resolve $\boldsymbol{u}_r$ on either side of the interface. In addition, the overlapping of snapshot sets tends to improve the accuracy of the transition, as will be shown in the results section.

5. Results

In this section we will discuss the results of three test cases to be introduced in what follows. The source code for these experiments is a custom finite volume incompressible flow solver that has been written in the C++ programming language using the Armadillo linear algebra library [89,90] and the LIS library of iterative solvers for linear systems [70]. All experiments in this paper have been carried out on a simple workstation with a single Intel(R) Core(TM) i7-6700HQ CPU.

5.1. Shear layer roll-up

5.1.1. Problem set-up

The first test case concerns the shear layer roll-up (SLR) [65,52], which is a flow on a double-periodic domain $[0, 2\pi] \times [0, 2\pi]$ with a central band of flow in a positive coordinate direction and neighbouring bands of flow in the opposite direction. The bands are joined with a thin region of strong velocity gradients and hence strong shear forces. The initial conditions are:

$$u(x, y, 0) = \begin{cases} \tanh\left(\frac{y - \pi/2}{\delta}\right), & y \leq \pi \\ \tanh\left(\frac{3\pi/2 - y}{\delta}\right), & y > \pi \end{cases}, \quad v(x, y, 0) = \epsilon \sin(x),$$
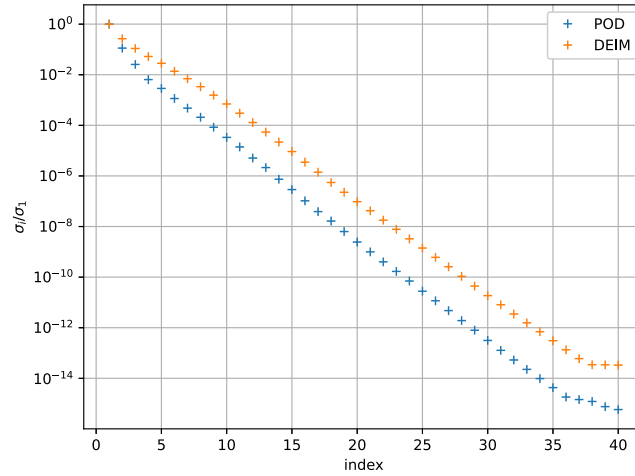
Fig. 4. The normalized POD and DEIM singular value decay of snapshot data from a FOM simulation with $\nu = 0$ and $t = 4$.

where $\delta = \frac{\pi}{15}$ determines the initial thickness of the shear layers and the parameter $\epsilon = 0.05$ determines the initial amplitude of an unstable perturbation in the second coordinate direction to trigger the so-called roll-up.

Using the SLR the proposed hROMs will be tested for their energy- and momentum-conserving capabilities, accuracy and computational performance. Energy and momentum conservation will be tested by considering the temporal evolution of errors in conserved quantities (momentum and kinetic energy) for the first four seconds of the inviscid SLR flow. Accuracy will be tested by considering two different errors, namely the error with respect to the FOM, which is defined as

$$\epsilon_u(t) = \left\| \boldsymbol{u}_h(t) - \boldsymbol{u}_r(t) \right\|_{\Omega_h},$$
(41)

and the best-approximation error

$$\epsilon_b(t) = \left\| (I - \Phi \Phi^T \Omega_h) \boldsymbol{u}_h(t) \right\|_{\Omega_h},$$
(42)

which forms a lower bound to the accuracy that can be obtained using r POD modes. Comparing ϵ_u and ϵ_b will give an indication on how close to optimal the hROM is. We will also analyse whether overfitting takes place by considering the temporal evolution of the Mahalanobis distance $d_M(\boldsymbol{c})$. To validate our choice of hyperparameter for the MCLSDEIM we will inspect a set of L-curves. Finally, the computational performance will be tested by measuring execution times associated to the offline and online phases of the hROMs and comparing these to the FOM.

Based on a grid convergence study it was concluded that a 256×256 grid and time step $\Delta t = 0.01$ were sufficiently fine to have a reliable FOM reference result, and these settings will be used in the numerical experiments unless otherwise mentioned. The same time step is also used to build the snapshot matrix and to simulate the ROMs. The hyperparameter of the MCLSDEIM will be set to $\alpha = 1 \cdot 10^{-14}$ and the hyperparameter of the OCLSDEIM will be set to $m_p = 2m$.

5.1.2. Conservation properties

The proposed hyper-reduction methods will be tested for their energy- and momentum-conserving properties. We expect K_r to be conserved for the inviscid case only when using the proposed hROMs in conjunction with energy-conserving Runge-Kutta (RK) methods (see Remark 1). Thus, the proposed methods and the conventional DEIM will be tested for the case $\nu = 0$ using the (implicit) energy-conserving Gauss-Legendre 4 (GL4) method [86], and compared also to the classical (explicit) Runge-Kutta 4 (RK4) method [88]. Time integration will take place until $t = 4$, since for longer time intervals numerical oscillations develop in the inviscid FOM simulation. These numerical oscillations do not destabilize the FOM solution due to the nonlinear stability property, but they render the simulation results inaccurate and corrupt the snapshot data sets and hence the ROM. Until $t = 4$ the FOM solution has a quickly decaying Kolmogorov N-width, and 8 POD and 8 DEIM modes suffice to accurately capture most of the energy in the snapshots as can be seen in Fig. 4. Note that the POD and DEIM modes are obtained from snapshots taken at every individual time step of the FOM simulation. Since the initial conditions of the SLR have a net total momentum of zero in all coordinate direction we will plot the absolute value of its components $|(\boldsymbol{P}_r)_i|$. We expect this value to remain zero up to machine precision for all hROMs. For a momentum-conserving semi-discretization any RK method trivially conserves momentum, hence we only use RK4 (and not GL4) in order to not clutter the results. For the kinetic energy we will plot the value:

$$\epsilon_K(t) := |K_r(t) - K_r(0)|,$$

which denotes the absolute deviation of K_r with respect to its initial value. As kinetic energy is conserved for $\nu = 0$ by the energy- and momentum-conserving hROMs, this value is also expected to remain zero up to machine precision using energy-conserving RK methods.
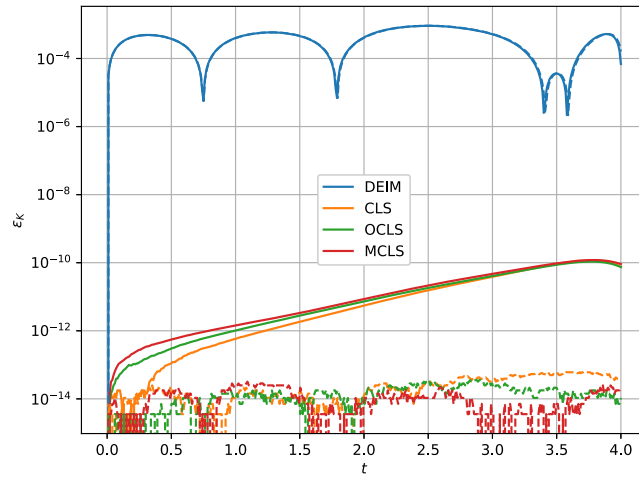
Fig. 5. The error ϵ_K in kinetic energy with respect to its initial value for the inviscid SLR using the proposed hROMs and the DEIM-hROM, the solid lines are for RK4 and the dashed lines for GL4. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)
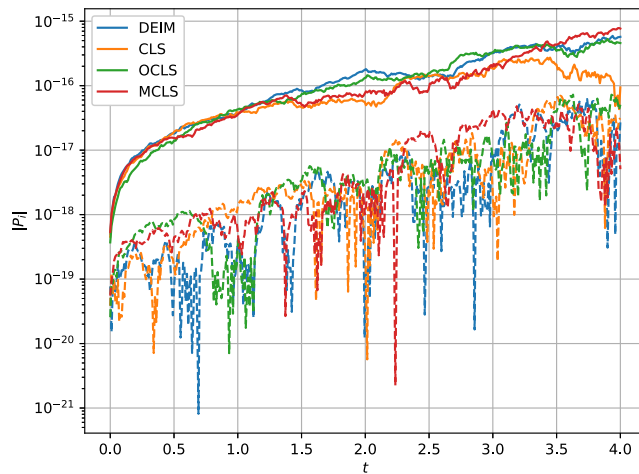

Fig. 6. Conservation of momentum component $(P_r)_i$ in absolute value for the inviscid SLR using the proposed hROMs and the DEIM-hROM and RK4, the solid lines are for $i = 1$ and the dashed lines are for $i = 2$.

In Fig. 5 the error in the kinetic energy is displayed. Colors are used to distinguish the different hROM methods and solid lines and dashed lines are used to distinguish RK4 from GL4 time integration. The key observation is that when using the standard DEIM algorithm (with either RK4 or GL4), K_r deviates significantly from its initial value, while *our proposed energy-conserving hROMs exactly conserve kinetic energy up to machine precision* (using GL4 time integration). When using the RK4 time integrator the energy-conserving methods deviate slightly from their initial value, due to the non-energy-conserving nature of RK4. These results are in line with [87] in which similar small energy errors were observed when using high-order explicit RK methods to integrate systems with energy-conserving semi-discretizations.

In Fig. 6 the error in the momentum components is displayed. In this figure colours are used to distinguish which hROM method is applied and solid and dashed lines are used to indicate the first and second component of P_r, respectively. It can be observed that the theoretical results on momentum conservation hold for all hROMs including the DEIM-hROM.

5.1.3. Convergence behaviour

To assess the behaviour of ϵ_u as r, the dimension of the POD space, and m, the dimension of the DEIM space, are varied we perform a convergence study. This will be done for the case of viscous flow, with $\nu = 0.001$, allowing integration of the SLR flow until $t = 8$. Given the small differences between RK4 and GL4 in the study of conservation properties, we only consider RK4 here for efficiency. Our primary interest is the behaviour of ϵ_u as a function of m, hence we will sweep this parameter for some values of r. To determine what values of m and r are necessary such that the corresponding reduced spaces $\mathcal{V}$ and $\mathcal{M}$ can accurately resolve the snapshot data we study again the singular value decay of X and Ξ. The singular value decays of the snapshot data for a FOM simulation with $\nu = 0.001$ and until $t = 8$ have been plotted in Fig. 7 for the data sets X and Ξ. From this figure we determine a
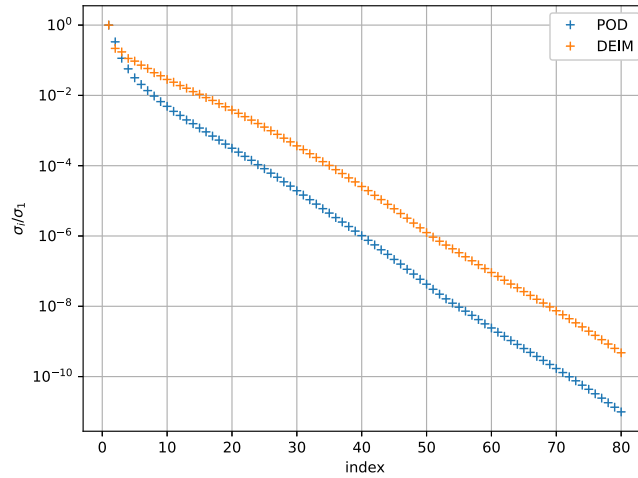
Fig. 7. The normalized POD and DEIM singular value decay of snapshot data from a FOM simulation with $v = 0.001$ and $t = 8$.
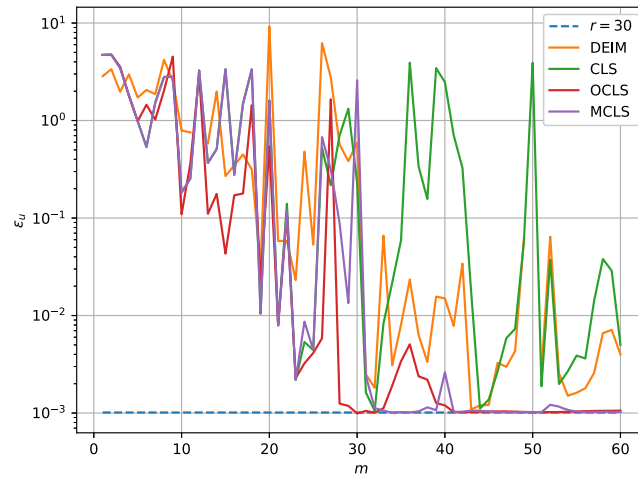

Fig. 8. Convergence study of ϵ_u evaluated at $t = 8$ for different hROMs and different values of m for $r = 30$, including the non-hyper-reduced model results (dashed).

sweep will be performed over all values $m \in \{1, 2, ..., 60\}$ corresponding to a relative information content (RIC) [55] ranging from 89.86% to 100%. The reduced space dimension will be taken as $r \in \{30, 40\}$ both corresponding to RIC $\approx 100\%$.

The results of the convergence study have been plotted in Fig. 8 and Fig. 9 for $r = 30$ and $r = 40$ respectively. In these figures we also plot ϵ_u for the non-hyper-reduced order model (dashed lines). This value will be referred to as the 'ROM error'. Clearly, the ROM error is independent of the dimension m of $\mathcal{M}$. Although the ROM error is not strictly a lower bound for ϵ_u of hROMs, the methods should converge to it if the convection operator is approximated accurately enough. All methods show erratic error behaviour as m increases for $m < r$. This shows that the convection operator cannot be expected to be approximated well throughout a long time-integration procedure when m is not sufficiently large compared to r. When $m \geq r$ better accuracy is obtained with the OCLSDEIM and MCLSDEIM, and the ROM error is closely approximated. This means that the convection operator is approximated well enough throughout the time integration such that $\Phi^T C_h(\Phi a)$ and $\Phi^T M c(a)$ are nearly indistinguishable. It can be seen that ϵ_u as obtained using the DEIM and CLSDEIM remains behaving erratically when $m \geq r$ and generally the ROM error is not obtained. This shows that exactly (DEIM) or near exactly (CLSDEIM) fitting the approximation $Mc(a)$ to $C_h(\Phi a)$ in $\mathcal{P}$ when $m_p = m$ is not sufficient to guarantee accuracy in a long time integration process. Indeed, using more measurements (OCLSDEIM) or smarter choices of DEIM coordinates (MCLSDEIM) is required to obtain the ROM error.

5.1.4. Temporal error evolution

To obtain a better understanding of the performance of the methods we will analyse the full temporal evolution of several errors. We will examine a particularly bad case for the DEIM and CLSDEIM such that the errors are further exaggerated, specifically we will look at $r = 30, m = 50$, for which Fig. 8 shows high errors for both DEIM and CLSDEIM. For illustration, the velocity and vorticity fields as obtained using $r = 30, m = 50$ are shown in Fig. 10. For this case the error due to the DEIM becomes unstable and due to the CLSDEIM becomes very large. Besides ϵ_u, we will also analyse the temporal evolution of the operator error:
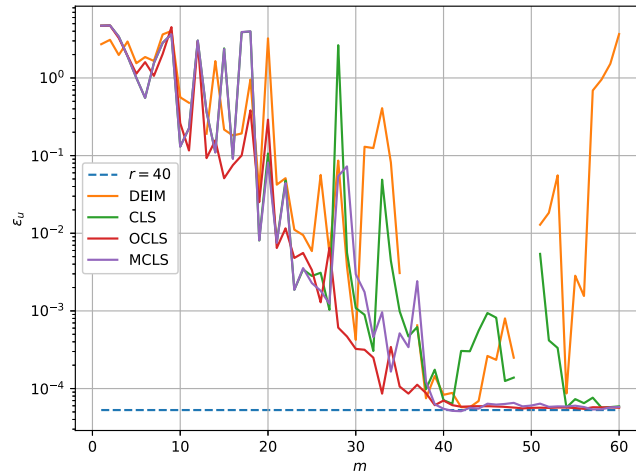
Fig. 9. Convergence study of ϵ_u evaluated at $t = 8$ for different hROMs and different values of m for $r = 40$, including the non-hyper-reduced model results (dashed).
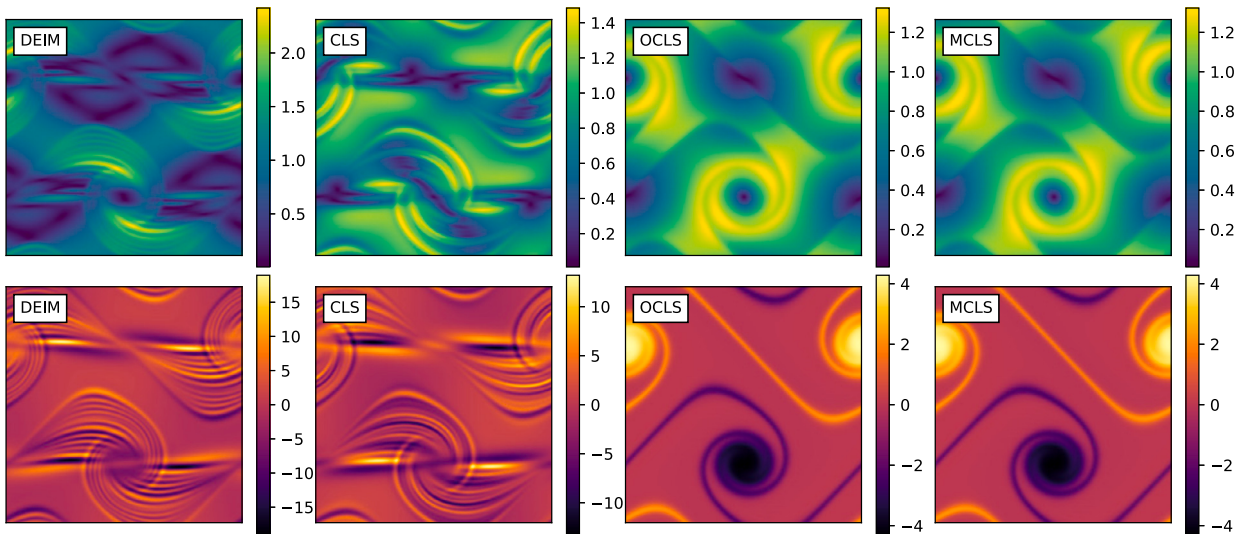

Fig. 10. The velocity and vorticity fields as predicted by the different hyper-reduction methods at $r = 30, m = 50, t = 8$ ($t = 5$ for the unstable DEIM).

$$\epsilon_M(t) := \frac{||C_h(\Phi a(t)) - M c(a(t))||_{\Omega_h}}{||C_h(\Phi a(t))||_{\Omega_h}},$$

and the operator projection error:

$$\epsilon_\pi(t) := \frac{||(I - M M^T) C_h(\tilde{u}_r(t))||_{\Omega_h}}{||C_h(\tilde{u}_r(t))||_{\Omega_h}},$$

where $\tilde{u}_r : \mathbb{R}^+ \to \mathbb{R}^N$ is the solution to the non-hyper-reduced ROM. These errors provide insight in how well a hyper-reduction method is approximating the convection operator.

In Fig. 11 the error evolution of both ϵ_u and ϵ_M have been displayed for $r = 30, m = 50$. The ideal projection error ϵ_b has also been plotted as function of time. We once more use colours to distinguish different methods, solid lines are associated to the evolution of ϵ_u using the RK4 integrator and dashed lines to ϵ_M also using the RK4 integrator. We can see that the errors due to DEIM and CLSDEIM grow gradually as time progresses. Furthermore, as the solution becomes more dominated by errors it becomes more difficult for the DEIM and CLSDEIM to accurately approximate the correct convection operator output. The OCLSDEIM and MCLSDEIM are much more accurate throughout the full time integration process. It can be seen that their respective ϵ_u values stay close to the lower bound ϵ_b and that the operator output is approximated very accurately.

Plotting the Mahalanobis distance $d_M(c(a))$ in Fig. 12, the poor accuracy of the DEIM coordinates as determined by the DEIM and CLSDEIM is accentuated. As time progresses the values of the DEIM coordinates becomes increasingly less similar to the reference values in $M^T \Xi$, whereas the DEIM coordinates as determined by the OCLSDEIM and MCLSDEIM remain close to the reference
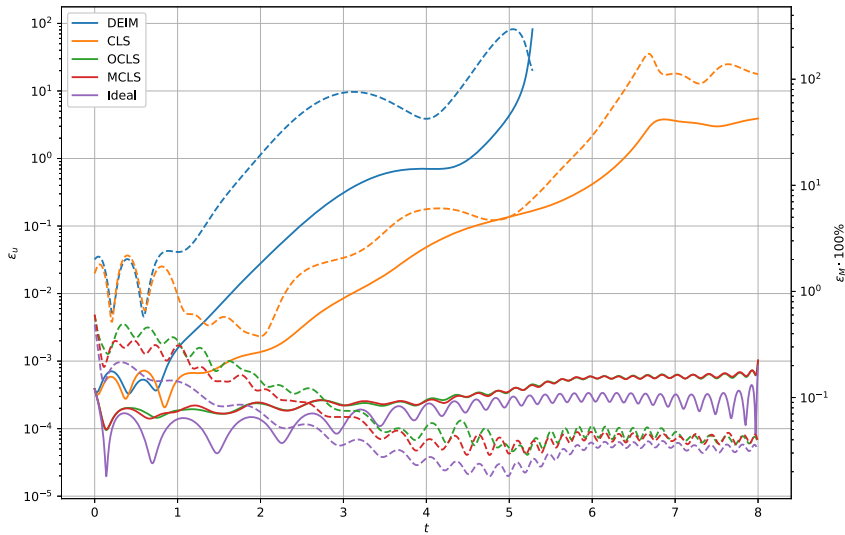
Fig. 11. The evolution of ϵ_u (solid lines) and ϵ_M (dashed lines) for $r = 30$ and $m = 50$ using different hROMs and using RK4, compared to the projection errors ϵ_b (purple solid line) and ϵ_π (purple dashed line).
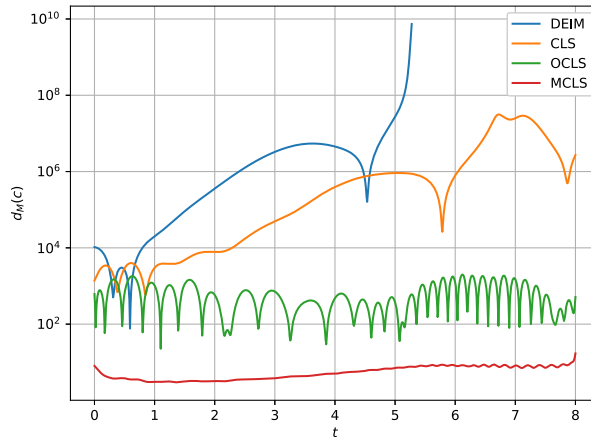

Fig. 12. Temporal evolution of the Mahalanobis distance $d_M(c)$ of the DEIM coordinates using $r = 30, m = 50$ for the different hyper-reduction methods.

distribution. Here, similar is in the sense of the Mahalanobis distance. Furthermore, Fig. 12 shows that the Mahalanobis distance, at least in the reproduction case, is a good indicator of a poor choice in DEIM coordinates as it becomes increasingly larger when ϵ_M becomes larger. Combining these results and the previous results on error evolution, we see that by exactly (DEIM) or near exactly (CLSDEIM) fitting the approximation $Mc(a)$ to $C_h(\Phi a)$ in $\mathcal{P}$ with $m_p = m$, DEIM coordinates are determined that cause gradual accumulation of errors in the time integration process. Instead, more robustness is required to deal with demanding situation like time integration where accumulation of errors can cause problems. As shown in our experiments, robustness of the DEIM and CLSDEIM can be significantly improved by:

1. performing a regression through a larger set of measurements than free parameters (OCLSDEIM);
2. using knowledge on the prior distribution of DEIM coordinates (MCLSDEIM).

5.1.5. MCLSDEIM hyperparameter validation

We will briefly validate the choice of $\alpha = 1 \cdot 10^{-14}$ for the MCLSDEIM (31) by means of L-curves. To do so, we will consider a single reduced space configuration $r = 30, m = 50$ and calculate the residual norm, $||P^T C_h(\Phi a) - P^T M c_\alpha||$, and the Mahalanobis distance, $d_M(c_\alpha) = ||c_\alpha - \mu_C||_{S^{-1}}$, for the MCLSDEIM coefficients c_α associated to hyperparameter values in the range $\alpha \in [1 \cdot 10^{-17}, 9 \cdot 10^{-10}]$ at all times $t_L \in \{1, 2, ..., 8\}$. We will calculate the required value of a with the accurate OCLSDEIM method also with $r = 30, m = 50$ and $m_p = 100$. As a result, we will have an L-curve for every point in time $t_L \in \{1, ..., 8\}$ as shown in Fig. 13. To validate the choice of $\alpha = 1 \cdot 10^{-14}$ we specifically mark the points on the L-curves in Fig. 13 with this hyperparameter value. By inspection, the dots are roughly in the corners of the L-curves (with exception of $t = 8$), indicating a good choice of hyperparameter. Note that in Fig. 13
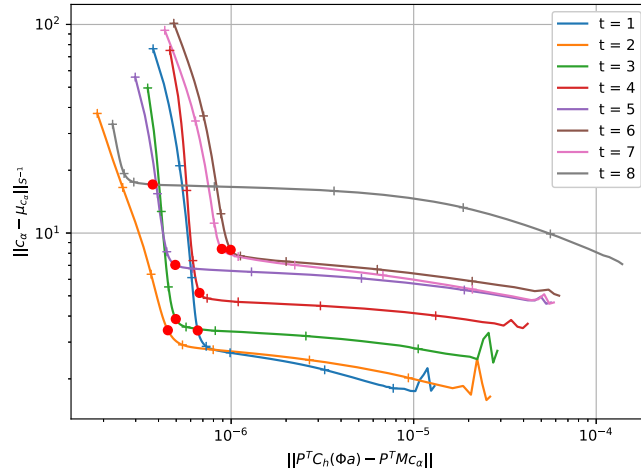
Fig. 13. L-curves at different moments in the SLR flow for $r = 30, m = 50, m_p = m$, the values associated to $\alpha = 1 \cdot 10^{-14}$ are marked with a red dot, the values associated to $\alpha \in \{10^{-17}, 10^{-16}, ..., 10^{-10}\} \setminus 10^{-14}$ have been marked with +-symbols and increase from left to right on the curves.

Table 1
Computational performance (offline and online) of the different hyper-reduction methods for $r = 30, m = 40$.

Method	Wall-clock time online (s)	Speed-up	Wall-clock time offline (s)
FOM	191	-	-
DEIM	0.236	808	45.105
CLSDEIM	0.300	636	45.104
OCLSDEIM(2m)	0.494	386	44.972
OCLSDEIM(4m)	1.099	174	45.123
MCLSDEIM	0.302	632	44.215

the ratio between the squared Mahalanobis distance and the squared residual norm is roughly $\mathcal{O}(10^{14})$ for $\alpha = 1 \cdot 10^{-14}$, hence in the MCLSDEIM loss-function (31) the terms $||P^T C_h(\Phi a) - P^T M c_\alpha||^2$ and $\alpha ||c_\alpha - \mu_C||^2_{S^{-1}}$ should be of the same order of magnitude. We will not repeat these experiments for the coming test-cases, but assume that if $\alpha = 1 \cdot 10^{-14}$ gives acceptable values for the error with respect to the FOM, ϵ_u, that the hyperparameter choice is good.

5.1.6. Performance

To end our discussion on the SLR experiment we will show that the hyper-reduction methods exhibit similar computational performance as the conventional DEIM. This will be done by measuring the on- and offline phase wall-clock times and the associated speed-ups with respect to the FOM. Here, the offline phase is constituted by the necessary SVDs to construct $\mathcal{M}$ and $\mathcal{V}$, the construction of $\mathcal{P}$ and the precomputation of the necessary matrices and LU-decompositions. The online phase is constituted by the numerical integration of the hROMs. We will evaluate this for $r = 30, m = 40$ as all methods provide accurate results for this configuration of reduced space dimensions. To see the effect of increasing m_p for the OCLSDEIM we will test for $m_p \in \{2m, 4m\}$, where we expect to see a decline in performance as m_p is increased.

The results are tabulated in Table 1 for the online and offline phase. It can be seen that the computational performance is similar for all methods, both in terms of online and offline performance. The DEIM, CLSDEIM and MCLSDEIM are especially close due to the fact that for all these cases $m_p = m$. The effect of the larger number of measurement evaluations required can be seen for the OCLSDEIM(2m) and OCLSDEIM(4m), which are slower than the other methods. Furthermore, using $m_p = 2m$ of the OCLSDEIM indeed provides better computational performance than $m_p = 4m$, as expected.

5.2. Two-dimensional homogeneous isotropic turbulence

5.2.1. Problem set-up

The second test case is two-dimensional homogeneous isotropic turbulence (2DT). 2DT is a flow on a double-periodic domain $[0, 2\pi] \times [0, 2\pi]$. We will generate initial conditions following the procedure outlined in [84]. In short, we construct an initial energy spectrum:

$$E(k) = \frac{a_s}{2} \frac{1}{k_p} \left(\frac{k}{k_p} \right)^{2s+1} \exp \left[-\left(s + \frac{1}{2} \right) \left(\frac{k}{k_p} \right)^2 \right],$$
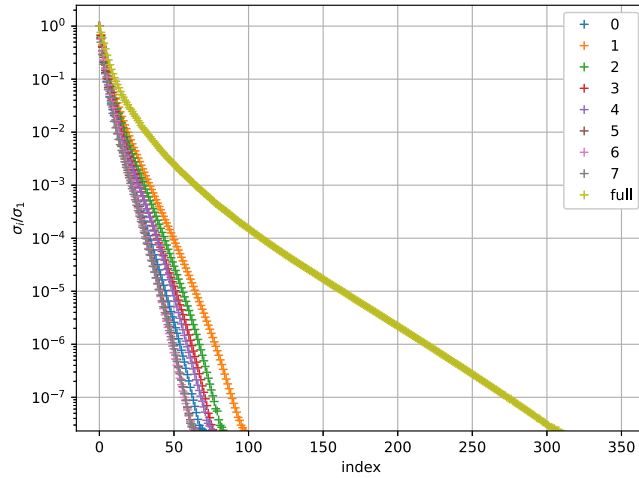
Fig. 14. Singular value decay of solution snapshots of intervals versus the full snapshot set.

where $k := ||\mathbf{k}|| = \sqrt{k_x^2 + k_y^2}$ is the wave-vector magnitude. The maximum value of the initial spectrum takes place at $k_p = 12$, $a_s \in \mathbb{R}$ is a normalization factor given by $a_s := (2s + 1)^{s+1}/(2^s s!)$ and $s \in \mathbb{R}$ is a shape parameter taken as $s = 3$. From this spectrum we follow [84] and calculate vorticity Fourier-coefficients like:

$$|\hat{\omega}(\mathbf{k})| = \sqrt{\frac{k}{\pi} E(k)}.$$

Consequently, we generate a vorticity spectrum $\hat{\omega} : \mathbb{R}^2 \to \mathbb{C}$ by providing random phases to the Fourier-coefficients using a process for which we refer to [84]. Finally, we obtain a divergence-free velocity field by solving the relation:

$$\Delta \psi = -\omega, \quad \mathbf{u} = \left[\frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x} \right]^T,$$

numerically. Note that this procedure is strictly for two-dimensional flows.

2DT is a fluid flow that has been studied extensively [3,28,84]. Furthermore, it is a highly convection-dominated flow with many convected spatial structures and as a result it has a slowly decaying Kolmogorov N-width. For this reason the purpose of this test case is two-fold. Namely, the complexity of the flow makes it an adequate test case to see how well our proposed hyper-reduction methods will perform for more turbulent flows of engineering interest. Additionally, the slow Kolmogorov N-width decay will allow us to compare the PID and our newly proposed energy- and momentum-conserving alternative, the SP-PID, in a meaningful setting.

The FOM simulations are performed with a 1024×1024 grid and a time step of $\Delta t = 0.0002$. These settings are based on the observation that these also led to converged results for similar spatially second-order energy-conserving schemes in [84]. Saving snapshots every time step for this test case is not feasible given the simple workstation used to perform all calculations. Hence, the simulation will be sampled such that the Nyquist-Shannon criterion is met [3]; this criterion is satisfied by sampling every $\Delta t_s = 0.005$.

5.2.2. Hyper-reduction methods

Due to the previous results for the SLR flow, in this section we will only test the two best-performing methods, MCLSDEIM and the OCLSDEIM. We simulate the FOM until $t = 4$ with $\nu = 0.001$ and compare how well the resulting energy spectra $E(k)$ of the FOM and hROMs align at $t \in \{1, 2, 3, 4\}$. We will also provide a visual comparison of the vorticity fields at these time instances. To deal with the slow Kolmogorov N-width decay we will apply the SP-PID with $n = 8$ evenly sized intervals of 100 snapshots and for all intervals we will take overlap parameters $\gamma_l = \gamma_r = \gamma = 20$. The singular values of all intervals and the full snapshot sets have been displayed in Fig. 14 and Fig. 15 for the POD and DEIM, respectively. As discussed in [3], it can be seen that the application of temporal localization indeed leads to a faster decay of singular values within intervals. This effectively increases the Kolmogorov N-width decay. For the reduced space dimensions we will use $r = 30$ and $m = 40$ for all intervals. Like [84], we will calculate the energy spectra using the following equations:

$$E(k, t) = \sum_{k \leq ||\tilde{\mathbf{k}}|| < k+1} \hat{E}\left(\tilde{\mathbf{k}}, t\right), \quad \hat{E}(\mathbf{k}, t) = \frac{1}{2} k^2 |\hat{\psi}(\mathbf{k}, t)|^2,$$

where $\hat{\psi} : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{C}$ are the Fourier-coefficients of the numerically calculated stream-function associated to a discrete velocity field $\mathbf{u}_h(t)$.

We have provided the energy spectra for the different hyper-reduction methods in Fig. 16 and Fig. 17. Excellent agreement at all points in time can be observed for both the hyper-reduction methods. Furthermore, the results are in line with those observed
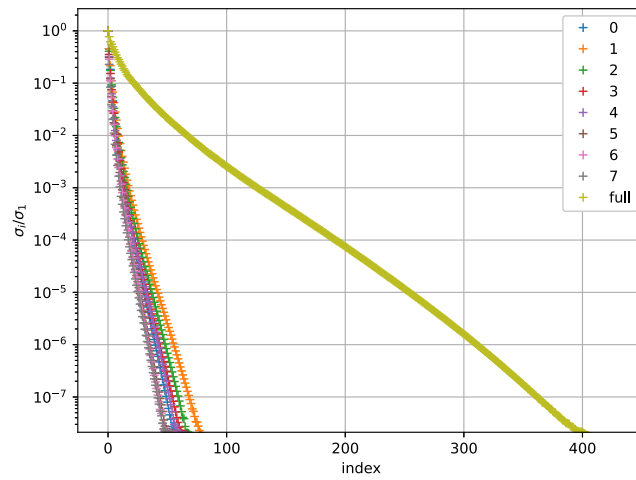
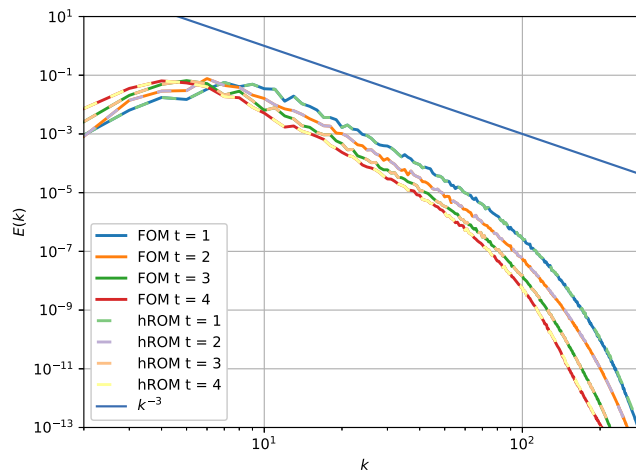**Fig. 15.** Singular value decay of operator snapshots of intervals versus the full snapshot set.

Fig. 16. Energy spectra at $t \in \{1, 2, 3, 4\}$ as calculated by the FOM (solid line) and MCLSDEIM-SP-PID hROM (dashed line).
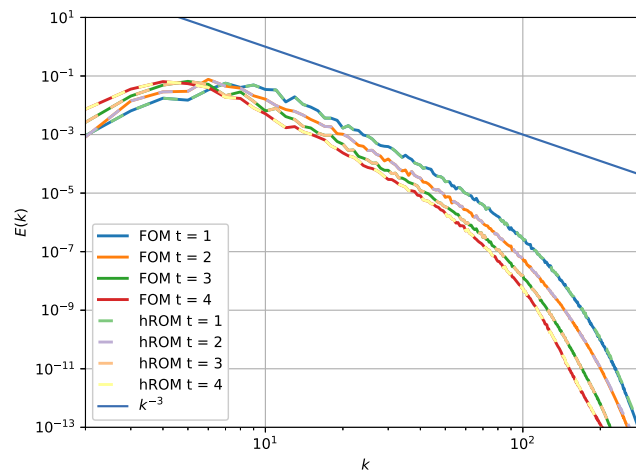

Fig. 17. Energy spectra at $t \in \{1, 2, 3, 4\}$ as calculated by the FOM (solid line) and OCLSDEIM-SP-PID hROM (dashed line).
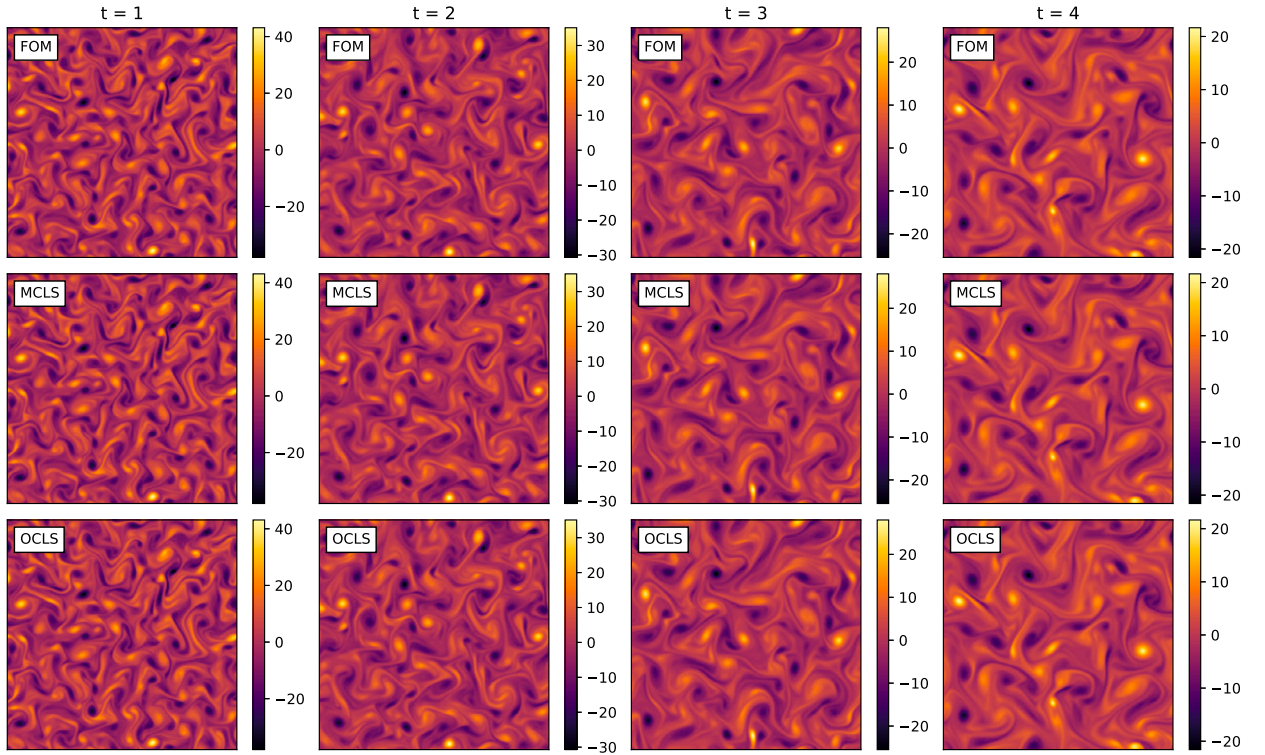
Fig. 18. 2DT vorticity fields as calculated using the FOM and hROMs for $t \in \{1, 2, 3, 4\}$.

in [84] for $Re = 1000$ in terms of spectra, where an accurate high-order pseudospectral method was used. We can see clearly that the spectrum is shifting to the left as time progresses, implying the presence of an inverse energy cascade. Furthermore, the inertial range nearly scales as $\mathcal{O}(k^{-3})$ which should be the case as $\nu \to 0$ [28]. We have also shown the vorticity fields as calculated using the FOM and the different hROMs in Fig. 18. We can see that, besides the energy spectra matching between the FOM and hROMs, the vorticity fields are also almost identical. From this we conclude that the hROMs using the MCLSDEIM and OCLSDEIM and the SP-PID:

1. can accurately reconstruct the FOM results;
2. adhere to the physics underlying the 2DT flow.

5.2.3. Performance

We will briefly summarize the computational performance of the hROMs compared to the FOM for the 2DT flow. The execution times for both the online as offline phases of the hROMs and the full execution time of the FOM have been tabulated in Table 2. As can be seen the FOM takes a very long time to complete. This can be attributed to the naive and serial implementation of the C++ code which is not optimized for relatively large two-dimensional problems. Furthermore, the wall-clock times of the offline phases for both hROMs were heavily influenced by the large amount of data associated to storing the snapshot matrices in memory while calculations were performed. Extrapolating computation times while having less memory occupied indicate that a wall-clock time of $\approx 85s$ for the offline phases of both MCLSDEIM as OCLSDEIM(2m) seems more accurate for a memory efficient implementation of the 2DT case.

5.2.4. Temporal localization methods

We will now study in more detail the conservation properties of SP-PID at interfaces and compare its transition accuracy against the orthogonal projection of the standard PID as in equation (36). Furthermore, we will briefly investigate the impact of the overlap parameter γ. We will do this by simulating the FOM until $t = 1$ with $\nu = 0.001$ and only considering $n = 2$ evenly sized intervals. We will then construct the reduced spaces associated to these intervals with overlap parameters $\gamma \in \{0, 20\}$ and analyse the differences this makes on the accuracy of the transition at $t = 0.5$. We will also consider the accuracy of the conserved quantities energy and momentum, which should be exactly conserved by the SP-PID and not by the PID. However, for the PID, we expect these quantities to be better conserved for larger γ. We also expect the accuracy of the transition to be improved by taking larger γ, for both SP-PID and PID. During this experiment we will use the MCLSDEIM with $\alpha = 1 \cdot 10^{-14}$ and $m = 40$. Additionally, we will use a reduced space of $r = 30$. Finally, we will determine the accuracy of the transition using the transition error:

Table 2

Computational performance (offline and online) of the different hyper-reduction methods for $r = 30, m = 40$.

Method	Wall-clock time online (s)	Speed-up	Wall-clock time offline (s)
FOM	58871.0	-	-
SP-PID-MCLSDEIM	7.985	7373	582.0
SP-PID-OCLSDEIM(2m)	15.172	3880	522.0

Table 3

Transition errors in velocity field and conserved quantities at the interface on $t = 0.5$.

	SP-PID ($\gamma = 0$)	PID ($\gamma = 0$)	SP-PID ($\gamma = 20$)	PID ($\gamma = 20$)
ϵ_T	0.0629	0.0629	0.00515	0.00515
$|K_r^+ - K_r^-|$	0	0.00198	0	$1.33 \cdot 10^{-5}$
$|(\boldsymbol{P}_r)_1^+ - (\boldsymbol{P}_r)_1^-|$	0	$2.83 \cdot 10^{-14}$	0	$3.32 \cdot 10^{-15}$
$|(\boldsymbol{P}_r)_2^+ - (\boldsymbol{P}_r)_2^-|$	0	$8.49 \cdot 10^{-17}$	0	$2.16 \cdot 10^{-16}$

$$\epsilon_T := ||\boldsymbol{u}_r^+ - \boldsymbol{u}_r^-||_{\Omega_h}.$$

The results of the experiment have been displayed in Table 3. It can be seen that the transition accuracy indeed increases as the overlap parameter is increased, since the transition error ϵ_T decreases significantly. Furthermore, ϵ_T using the SP-PID and the PID is nearly identical for both overlap parameter values. This shows that even in the unconstrained setting the SP-PID interface condition (38) is nearly optimal. As expected, the conserved quantities are exactly conserved over interfaces using the SP-PID. Increasingly smaller errors are made using the standard PID as the overlap parameter increases. We have also shown the energy spectra of the velocity fields before ($\boldsymbol{u}_r^-$) and after ($\boldsymbol{u}_r^+$) the interface in Fig. 19 and Fig. 20 for $\gamma = 0$ and $\gamma = 20$, respectively. We see good agreement between the new and old spectra for both methods, however for larger overlap parameters the smaller scales in the flow are captured more accurately. This can be seen in the zoomed Fig. 21 and Fig. 22 for $\gamma = 0$ and $\gamma = 20$, respectively. From this we conclude that:

1. the SP-PID exactly conserves both energy and momentum over interfaces, whereas the PID does not;
2. the overall accuracy of the transition (conserved quantities and ϵ_T) improves with greater overlap parameters γ for the SP-PID and PID.

We suspect there will be an optimal overlap parameter, this will be the subject of future research.

5.3. Vortex in uniform background flow

5.3.1. Problem set-up

The third test case is a vortex in a uniform background flow adapted from [52]. We will refer to the flow as AV for advected vortex. The flow is inviscid, i.e. $\nu = 0$, and takes place on a double periodic domain $[-4, 4] \times [-4, 4]$, so that kinetic energy will be conserved. The initial velocity field is given as:

$$u(x, y, 0) = 1.0 - yC \exp\left(-\frac{1}{2}r(x, y)^2\right), \quad v(x, y, 0) = xC \exp\left(-\frac{1}{2}r(x, y)^2\right),$$

with $r(x, y) = \sqrt{x^2 + y^2}$ and $C = 0.6944$. The initial velocity and vorticity fields have been displayed in Fig. 23 and Fig. 24 respectively, for later reference. The inviscid background flow will advect the vortex to the right without deformation or decay. The centre of the vortex will return at its initial position $\boldsymbol{x}_c = [0, 0]^T$ at $t = 8$ after which the process will repeat.

Its time-periodic nature makes AV an ideal flow to test the stability of the proposed hyper-reduction methods and to compare them to the standard DEIM when integrating over long time intervals. Namely, if the training data contains at least one period of the flow we in principle have captured all states the flow takes as $t \to \infty$. So, ideally, capturing one period of training data is equivalent to capturing the flow indefinitely in time. In practice, the FOM will contain some numerical error and hence will not be perfectly time-periodic, but nonetheless we expect the obtained POD and DEIM bases Φ and M obtained from one period of data to be accurate in terms of ideal projection errors for all t.

We will integrate the hROMs for several reduced space sizes and beyond the training horizon of $T = 8$ to show that energy-conserving methods have superior stability properties when integrated for long times compared to non-energy-conserving methods. Stability will be tested by monitoring the kinetic energy error $\epsilon_K(t)$ which should remain constant to machine precision for fully-discretely energy-conserving methods while it can increase arbitrarily upon instability. Although the time-periodic nature of the flow is ideal to test long time stability of the hROMs, we note that AV is actually not well-suited for linear model reduction methods like those proposed in this article. This is because AV is a pure convection problem and therefore it has very slow Kolmogorov N-width decay. We will deal with this problem by applying the SP-PID, taking into account the time-periodicity of AV.

The FOM simulations are performed on a 64×64 grid and with a time step of $\Delta t = 0.01$. These settings do not results in numerical wiggles and are sufficient to give accurate results. The same time step is used to build the snapshot matrices and to simulate the hROMs.
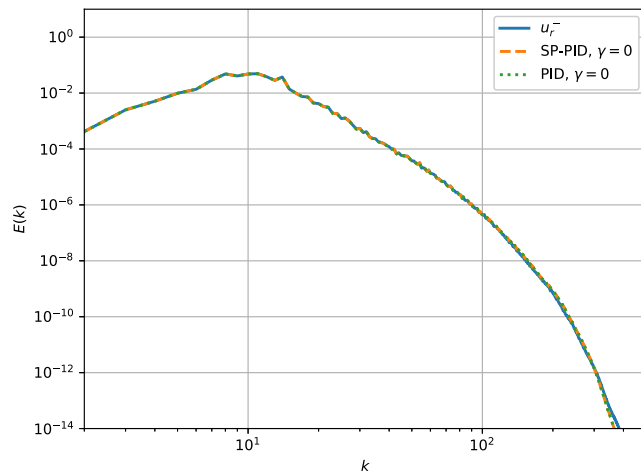
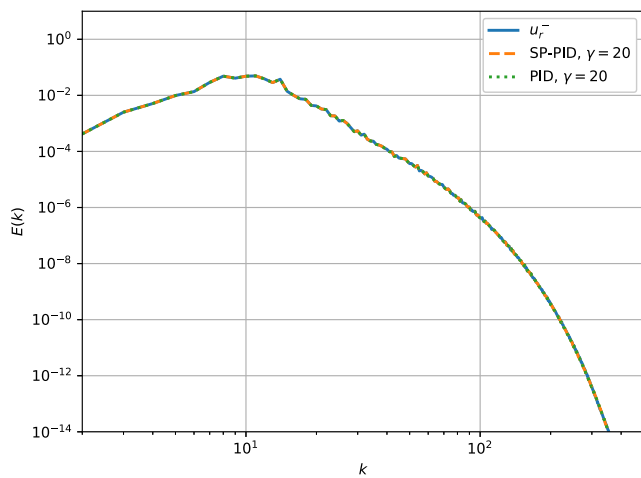**Fig. 19.** Spectra before and after transition at $t = 0.5$ using PID and SP-PID for $\gamma = 0$.

Fig. 20. Spectra before and after transition at $t = 0.5$ using PID and SP-PID for $\gamma = 20$.
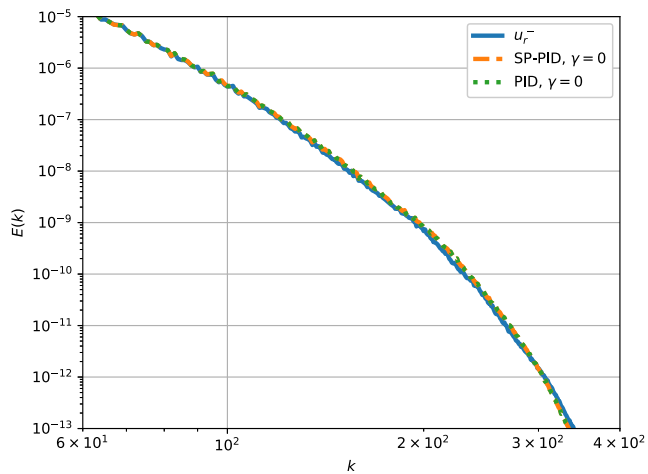

Fig. 21. Spectra before and after transition at $t = 0.5$ using PID and SP-PID for $\gamma = 0$ (zoomed).
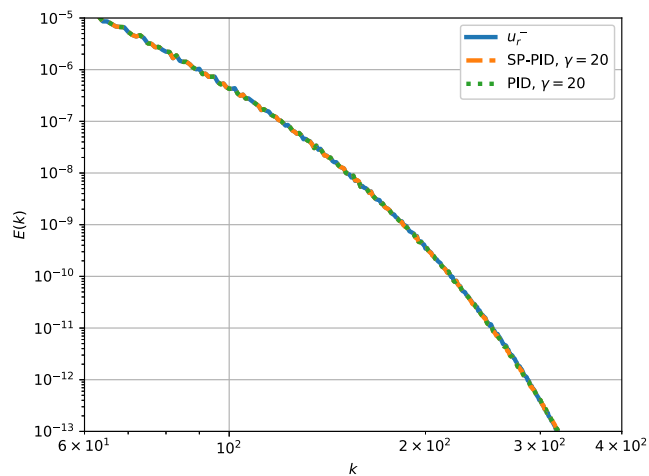
Fig. 22. Spectra before and after transition at $t = 0.5$ using PID and SP-PID for $\gamma = 20$ (zoomed).
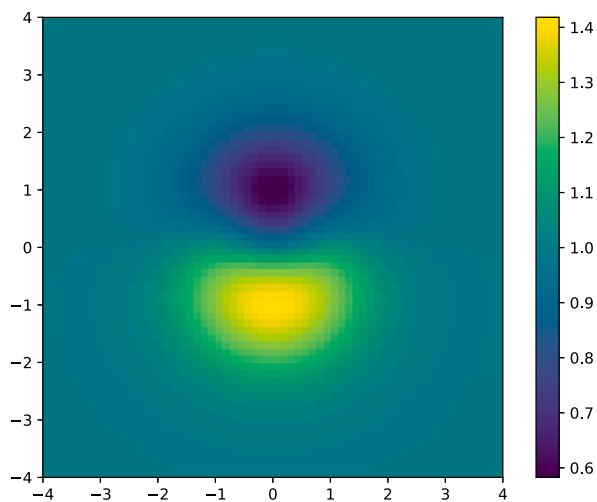

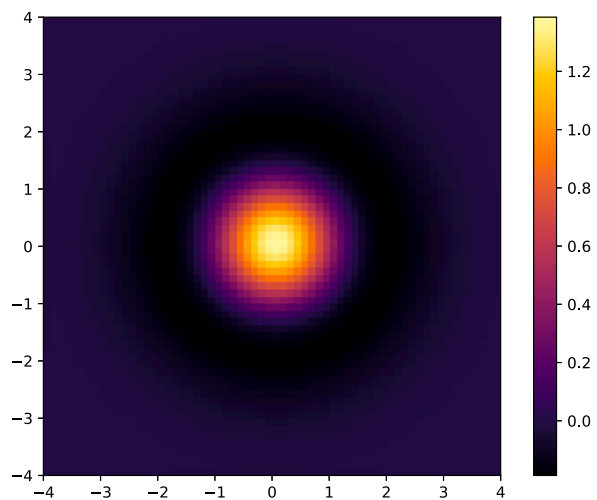**Fig. 23.** Velocity magnitude AV initial condition.

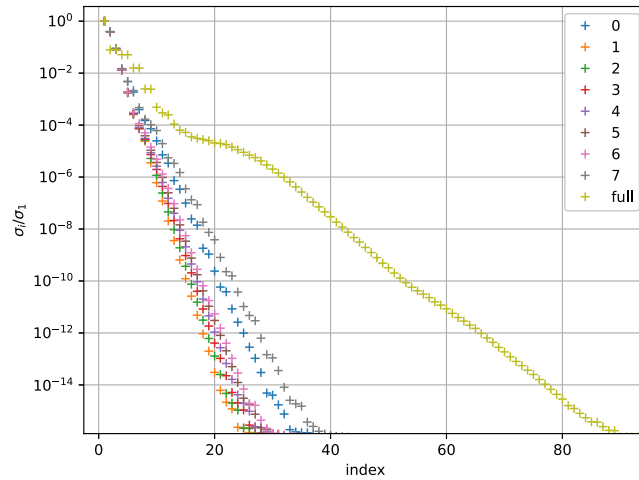**Fig. 24.** Vorticity field AV initial condition.

Fig. 25. Singular value decay of solution snapshots of intervals versus the full snapshot set.
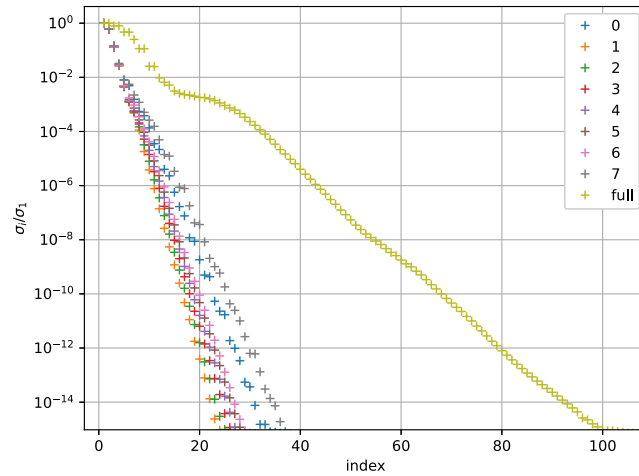

Fig. 26. Singular value decay of operator snapshots of intervals versus the full snapshot set.

5.3.2. Long time stability

We will first study the long time stability of the proposed energy-conserving methods by comparing them with non-energy-conserving methods. To deal with the slow Kolmogorov N-width decay we will apply the SP-PID with $n = 8$ evenly sized intervals of 100 snapshots and for all intervals we will take overlap parameters $\gamma_l = \gamma_r = \gamma = 20$. Note that we will also take an overlap between the first and last intervals as the system is time-periodic. The singular values of all intervals and the full snapshot sets have been displayed in Fig. 25 and Fig. 26 for the POD and DEIM, respectively. The experiment will be performed for two configuration of reduced spaces, the first configuration is $r = 15$ and $m = 20$ for all intervals, the second is $r = 20$ and $m = 30$ for all intervals. Both configurations have RIC $\approx 100\%$ for both POD and DEIM spaces and for every interval. We will use oversampling to improve the accuracy of both the non-energy-conserving as well as the energy-conserving hyper-reduction methods, i.e. oversampled DEIM (ODEIM) [75] and OCLSDEIM, respectively. Precisely, we will take $m_p = 100$ for both reduced space configurations and all intervals. Randomly placing measurement points for this flow is not sufficient since most points outside the vortex hardly experience any temporal changes and therefore contribute no information to the time evolution of the system at other places. Therefore, we will determine all measurement points using the DEIM algorithm [24] and $m = 100$ and after constructing $\mathcal{P}$ for all intervals truncate the DEIM space back to either $m = 20, 30$. We will integrate the hROMs until $t = 10T = 80$, far beyond the time for which the hROMs are trained. We will perform the experiment with both RK4 and GL4.

The resulting energy conservation error ϵ_K as a function of time has been displayed in Fig. 27. Failure due to instability can be observed for the non-energy-conserving ODEIM for all testing configurations. While in the training domain the ODEIM has $\epsilon_K \approx \mathcal{O}(10^{-7})$, the error sharply increases when leaving the training domain and then continues to accumulate while integrating in time until it fails catastrophically. For RK4 the kinetic energy becomes infinite, while for GL4 the Newton iterations fail to converge. It can be seen that for both RK4 and GL4 the energy-conserving hyper-reduction methods remain stable, with small deviations from energy conservation for RK4 and exact energy conservation up to machine precision for GL4. The hROMs do not only remain stable
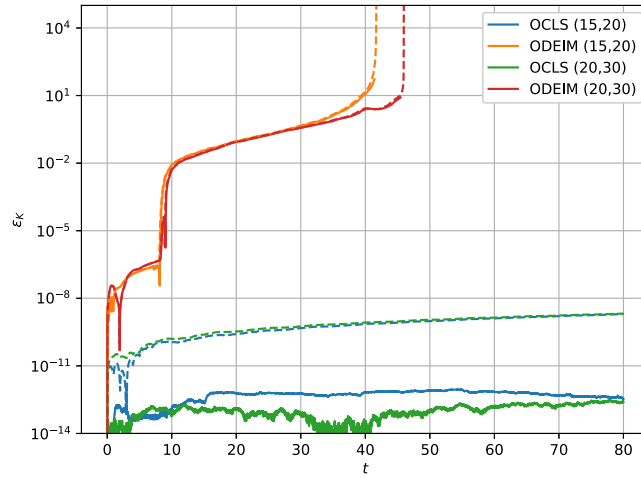
Fig. 27. The error ϵ_K in kinetic energy with respect to its initial value for AV using the proposed hROMs and the ODEIM-hROM and the different reduced space configurations indicated as (r, m). Note that results using RK4 are given with <u>dashed</u> lines and GL4 with <u>solid</u> lines.
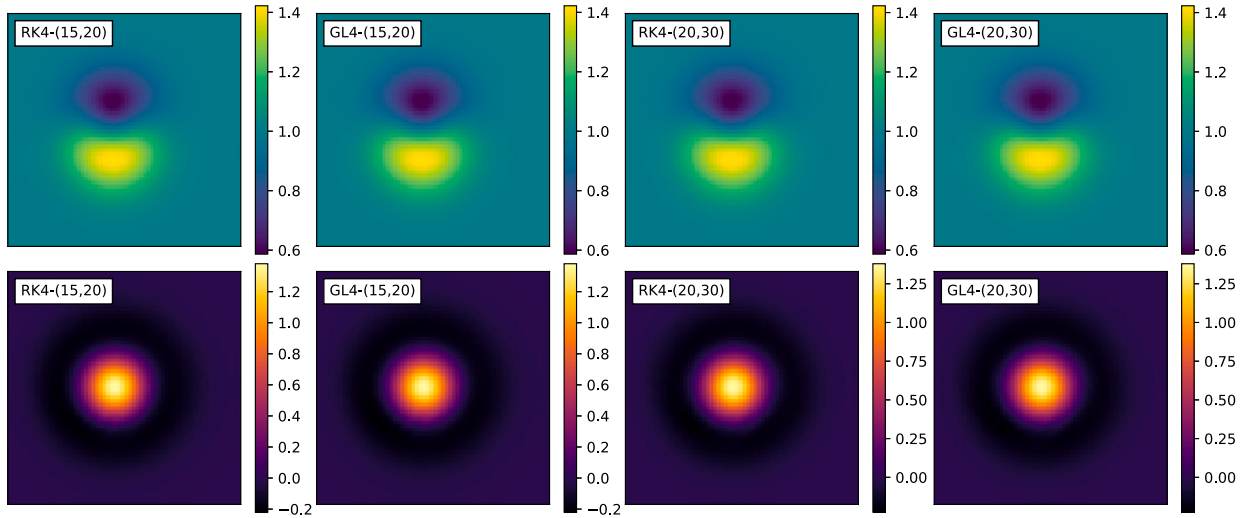

Fig. 28. The velocity and vorticity fields of AV at $t = 80$ calculated using the OCLSDEIM using different reduced space configurations indicated as (r, m) and time integrators.

but also remain very accurate as can be seen in Fig. 28, suffering only of a slight phase error compared to Fig. 23 and Fig. 24. These results strongly highlight the importance of energy conservation, especially when it is desired to integrate the hROMs for longer periods in time, and they stress that oversampling alone is not sufficient to guarantee stability.

5.3.3. Performance

We will briefly summarize the computational performance of the hROMs compared to the FOM for one period of the AV flow. The execution times of the online and offline phases of the hROMs and the execution time of the FOM, all for one AV period, have been tabulated in Table 4. Significant speed ups can be observed for the hROMs using the explicit RK4 time integrator even while the FOM is computed on a rather coarse 64×64 computational mesh. All offline phases are comparable in duration, with those of the ODEIM being slightly shorter because less operator precomputation is necessary. The hROMs using the GL4 time integrator take significantly longer to finish due to the fact that a computationally expensive Newton iteration procedure is carried out every timestep. The execution times of the hROMs using the RK4 time integrator scale weakly with the reduced space configurations, while the execution times of the hROMs using GL4 scale relatively strongly. This scaling is a result of the Jacobian evaluations and inversions which are necessary for the Newton iterations used during the GL4 time integration. These performance results suggest the use of Jacobian-free nonlinear solvers in the future [58,53], if implicit time integration is desired.

Table 4

Computational performance (offline and online) of the different hyper-reduction methods and different reduced space configurations.

Method	Wall-clock time online (s)	Speed-up	Wall-clock time offline (s)
FOM	153.5	-	-
OCLS RK4 (15,20)	0.234	656	4.936
OCLS RK4 (20,30)	0.256	600	5.317
ODEIM RK4 (15,20)	0.239	642	4.352
ODEIM RK4 (20,30)	0.275	558	4.344
OCLS GL4 (15,20)	7.534	20.4	5.259
OCLS GL4 (20,30)	8.795	17.5	5.699
ODEIM GL4 (15,20)	7.128	21.5	4.704
ODEIM GL4 (20,30)	10.360	14.8	4.365

6. Conclusion

In this article we have proposed new energy- and momentum-conserving hyper-reduction methods referred to as the CLSDEIM, OCLSDEIM and MCLSDEIM. In these new hyper-reduction methods the conventional DEIM is posed as a minimization problem and subsequently constrained to conserve kinetic energy, in addition to conserving momentum. The OCLSDEIM and MCLSDEIM enhance the robustness of the CLSDEIM by oversampling and by Mahalanobis regularization, respectively.

The second novelty in this work is that we have proposed a method to conserve energy and momentum over interfaces in order to use the new hyper-reduction methods in combination with the PID temporal localization method. The method is based on solving a constrained minimization problem at every interface, and its solution can be interpreted as a projection followed by a scaling to conserve energy. This new SP-PID allows the construction of fully energy- and momentum-conserving and temporally localized hROMs of convection-dominated flows.

We have performed three numerical experiments to show the performance of the new energy- and momentum-conserving hyper-reduction methods and the energy- and momentum-conserving temporal localization method. We used the shear layer roll-up (SLR) to test the energy- and momentum-conserving properties, accuracy and computational performance of the proposed hyper-reduction methods. We confirmed the theoretical results that the CLSDEIM, OCLSDEIM and MCLSDEIM conserve energy and momentum in a fully discrete setting when the Gauss-Legendre 4 (GL4) Runge-Kutta (RK) time integration method is used. It was also shown that the standard DEIM caused a significant error in kinetic energy in this case. Convergence experiments showed that both the DEIM and CLSDEIM exhibit erratic behaviour as a function of the DEIM space dimension m, due to error accumulation during the time integration process. This erratic behaviour and error accumulation is alleviated by the robust OCLSDEIM and MCLSDEIM. Finally, the computational performance of the DEIM, CLSDEIM and MCLSDEIM was comparable, whereas the OCLSDEIM was slower than the other methods. This shows the benefit of using regularization instead of oversampling to increase robustness of the CLSDEIM.

The second numerical experiment was the two-dimensional homogeneous isotropic turbulence (2DT). The results of this test case further highlighted the accuracy and robustness of the new hyper-reduction methods. Furthermore, the 2DT experiment showed that using the SP-PID energy and momentum can be conserved over interfaces, in contrast to the classical PID. Finally, it was shown that overlapping snapshots in the construction of local reduced spaces can increase the transition accuracy.

Finally, the third experiment was an advected vortex in a uniform background flow (AV). Due to its time-periodic nature, AV was suitable to show the superior stability of energy-conserving methods compared to non-energy-conserving methods like the oversampled DEIM (ODEIM), especially when integrating for long periods of time. The ODEIM failed due to instability after a few periods, while the energy-conserving OCLSDEIM not only remained stable but also very accurate. Furthermore, the hROMs using the RK4 time integration method were significantly faster than those using the GL4 method.

There are several directions of future research that we believe to be valuable or interesting. First of all, we have mainly concerned ourselves with simple, two-dimensional test cases due to hardware limitations. We strongly believe that the proposed methods should also be tested on larger scale test cases like those in [10,39,38]. We suggest this as future research. Second, is the generalization of the MCLSDEIM and SP-PID to parametric problems. For the MCLSDEIM this entails confirming that the DEIM coordinates of a parametric problem can indeed be thought of as originating from a single prior distribution. For the SP-PID this entails finding methods to efficiently and accurately distribute the full parametric snapshot set over multiple subsets. Besides these methods, the extension to arbitrary boundary conditions (instead of periodic conditions) will be very valuable. Third, the methods proposed in this article could be merged with ideas from e.g. [23,97] to obtain energy- and momentum-conserving hROMs for problems with general first integrals *and* general nonlinear skew-symmetric operators (like the convection operator in this article or those in [66]). We believe this could be done by suitably altering condition (24) to conserve a general hyper-reduced first integral. Finally, we also suggest generalizing the norms used in the minimization problems of the proposed methods. For instance, using $(P^T M)^{-T}(P^T M)^{-1}$ or the inverse covariance matrix of $P^T C_h(\boldsymbol{u}_h(t))$ may prove useful.

CRediT authorship contribution statement

R.B. Klein: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **B. Sanderse:** Conceptualization, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Well-posedness of CLSDEIM minimization problem

We will show the existence and uniqueness of solutions to the CLSDEIM minimization problem (25). The CLSDEIM minimization problem:

$$\arg\min_{\boldsymbol{c}\in\mathbb{R}^m}\left|\left|P^T C_h(\Phi\boldsymbol{a}) - P^T M \boldsymbol{c}\right|\right|^2 \quad \text{s.t.} \quad \boldsymbol{a}^T\Phi^T M \boldsymbol{c} = 0,$$

can be rewritten by using the fact that $P^T M$ is nonsingular, as shown in [24]. We define:

$$\boldsymbol{y} := P^T M \boldsymbol{c},$$

and since $P^T M$ is nonsingular we can rewrite the minimization problem to one in terms of $\boldsymbol{y}$:

$$\arg\min_{\boldsymbol{y}\in\mathbb{R}^m}\left|\left|P^T C_h(\Phi\boldsymbol{a}) - \boldsymbol{y}\right|\right|^2 \quad \text{s.t.} \quad \boldsymbol{a}^T\Phi^T M (P^T M)^{-1}\boldsymbol{y} = 0.$$

The feasible set $\mathcal{F}_y(\boldsymbol{a}) \subseteq \mathbb{R}^m$ of this minimization problem is:

$$\mathcal{F}_y(\boldsymbol{a}) = \ker\left(\left\langle (M^T P)^{-1} M^T \Phi\boldsymbol{a}, \cdot\right\rangle\right) \subseteq \mathbb{R}^m.$$

Clearly, $\mathcal{F}_y(\boldsymbol{a})$ is a convex subset of $\mathbb{R}^m$ since for any $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{F}_y(\boldsymbol{a})$ and any $\theta \in [0,1]$ we have:

$$\left\langle (M^T P)^{-1} M^T \Phi\boldsymbol{a}, \theta\boldsymbol{y}_1 + (1-\theta)\boldsymbol{y}_2\right\rangle = \boldsymbol{a}^T\Phi^T M (P^T M)^{-1}(\theta\boldsymbol{y}_1 + (1-\theta)\boldsymbol{y}_2)$$

$$= \theta\boldsymbol{a}^T\Phi^T M (P^T M)^{-1}\boldsymbol{y}_1 + (1-\theta)\boldsymbol{a}^T\Phi^T M (P^T M)^{-1}\boldsymbol{y}_2$$

$$= 0,$$

and thus that $\theta\boldsymbol{y}_1 + (1-\theta)\boldsymbol{y}_2 \in \mathcal{F}_y(\boldsymbol{a})$. This holds even when $\boldsymbol{a} \in \ker(M^T\Phi)$, in this case $\mathcal{F}_y(\boldsymbol{a}) = \mathbb{R}^m$. Moreover, since the norm is a nonnegative quantity and taking the square of a nonnegative number is a convex operation, solving the CLSDEIM minimization problem is equivalent to minimizing:

$$\delta = \inf_{\widetilde{\boldsymbol{y}}\in\mathcal{F}_y(\boldsymbol{a})}||P^T C_h(\Phi\boldsymbol{a}) - \widetilde{\boldsymbol{y}}||.$$

Showing that there exists a unique minimizer $\boldsymbol{y} \in \mathcal{F}_y(\boldsymbol{a})$, with $\mathcal{F}_y(\boldsymbol{a})$ a convex subset of the Hilbert space $(\mathbb{R}^m, \langle\cdot,\cdot\rangle)$, such that $\delta = ||P^T C_h(\Phi\boldsymbol{a}) - \boldsymbol{y}||$ is standard, a proof can be found in [54, theorem 3.3-1]. By invertibility of $P^T M$ the solution to (25) therefore also exists and is unique.

Appendix B. CLSDEIM is an oblique projection

We will show that the CLSDEIM operator $\Pi_{\mathcal{M}_\mathcal{F}} = M\boldsymbol{c}(\boldsymbol{a})$, with $\boldsymbol{c}(\boldsymbol{a})$ as in (28), is an oblique projector, i.e. it is a non-symmetric, idempotent, linear operator. We will use the conveniently defined expression in (27) for the proof:

$$\Pi_{\mathcal{M}_\mathcal{F}}\boldsymbol{u} = M\left(A^{-1} - A^{-1}B(B^T A^{-1}B)^{-1}B^T A^{-1}\right)2(P^T M)^T P^T\boldsymbol{u}.$$

Clearly, $\Pi_{\mathcal{M}_\mathcal{F}}$ is a linear operator. The composition $\Pi_{\mathcal{M}_\mathcal{F}} \circ \Pi_{\mathcal{M}_\mathcal{F}}\boldsymbol{u}$ is given as:

$$\Pi_{\mathcal{M}_\mathcal{F}} \circ \Pi_{\mathcal{M}_\mathcal{F}} \boldsymbol{u} = M \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) 2 (P^T M)^T P^T M$$
$$\left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) 2 (P^T M)^T P^T \boldsymbol{u}$$
$$= M \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) A \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) 2 (P^T M)^T P^T \boldsymbol{u}$$
$$= M \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) \left(I - B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) 2 (P^T M)^T P^T \boldsymbol{u}$$
$$= \Pi_{\mathcal{M}_\mathcal{F}} \boldsymbol{u} - M \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) B (B^T A^{-1} B)^{-1} B^T A^{-1} 2 (P^T M)^T P^T \boldsymbol{u}$$
$$= \Pi_{\mathcal{M}_\mathcal{F}} \boldsymbol{u} - M \left(A^{-1} B (B^T A^{-1} B)^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} \right) B^T A^{-1} 2 (P^T M)^T P^T \boldsymbol{u}$$
$$= \Pi_{\mathcal{M}_\mathcal{F}} \boldsymbol{u},$$

making it idempotent. Taking the transpose of the projector $\Pi_{\mathcal{M}_\mathcal{F}}$ we have:

$$(\Pi_{\mathcal{M}_\mathcal{F}})^T = (M \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) 2 (P^T M)^T P^T)^T$$
$$= 2 P P^T M \left(A^{-1} - A^{-1} B (B^T A^{-1} B)^{-1} B^T A^{-1} \right) M^T$$
$$\neq \Pi_{\mathcal{M}_\mathcal{F}},$$

therefore the CLSDEIM is an oblique projection.

Appendix C. Proof of error bounds on the CLSDEIM projection

We will provide a proof of the following error bound:

$$||C_h(\Phi \boldsymbol{a}) - M c(\boldsymbol{a})|| \leq ||(P^T M)^{-1}|| \cdot \left(||(I - M M^T) C_h(\Phi \boldsymbol{a})|| + ||(I - M K_M K_M^T M^T) C_h(\Phi \boldsymbol{a})|| \right)$$

where $c(\boldsymbol{a})$ is given by (28) and $\mathrm{span}(M K_M) = \mathcal{M}_\mathcal{F} \subset \mathcal{M}$ with $K_M \in \mathbb{R}^{m \times m-1}$ unless $\Phi \boldsymbol{a} \perp \mathcal{M}$ in which case $I = K_M \in \mathbb{R}^{m \times m}$. In what follows we will ignore the latter case, since then the energy conservation condition (24) is trivially satisfied. To start, we write (28), the CLSDEIM approximation of $C_h(\Phi \boldsymbol{a})$, in the convenient form:

$$M c(\boldsymbol{a}) = M (P^T M)^{-1} P^T C_h(\Phi \boldsymbol{a}) - \frac{\left\langle P^T C_h(\Phi \boldsymbol{a}), (M^T P)^{-1} M^T \Phi \boldsymbol{a} \right\rangle}{||(M^T P)^{-1} M^T \Phi \boldsymbol{a}||^2} M (P^T M)^{-1} (M^T P)^{-1} M^T \Phi \boldsymbol{a}.$$

Denoting the DEIM error by $\varepsilon_{\mathrm{DEIM}} := C_h(\Phi \boldsymbol{a}) - M (P^T M)^{-1} P^T C_h(\Phi \boldsymbol{a})$, we first express $\varepsilon_{\mathrm{CLS}} := C_h(\Phi \boldsymbol{a}) - M c(\boldsymbol{a})$ in terms of $\varepsilon_{\mathrm{DEIM}}$:

$$\varepsilon_{\mathrm{CLS}} = \varepsilon_{\mathrm{DEIM}} - \frac{\left\langle P^T C_h(\Phi \boldsymbol{a}), (M^T P)^{-1} M^T \Phi \boldsymbol{a} \right\rangle}{||(M^T P)^{-1} M^T \Phi \boldsymbol{a}||^2} M (P^T M)^{-1} (M^T P)^{-1} M^T \Phi \boldsymbol{a},$$

then use the triangle inequality for the norm of the error:

$$||\varepsilon_{\mathrm{CLS}}|| \leq ||\varepsilon_{\mathrm{DEIM}}|| + \frac{| \left\langle P^T C_h(\Phi \boldsymbol{a}), (M^T P)^{-1} M^T \Phi \boldsymbol{a} \right\rangle |}{||(M^T P)^{-1} M^T \Phi \boldsymbol{a}||} \cdot \left|\left| M (P^T M)^{-1} \frac{(M^T P)^{-1} M^T \Phi \boldsymbol{a}}{||(M^T P)^{-1} M^T \Phi \boldsymbol{a}||} \right|\right|.$$

M is orthogonal and we are multiplying with a unit vector, therefore we have:

$$||\varepsilon_{\mathrm{CLS}}|| \leq ||\varepsilon_{\mathrm{DEIM}}|| + \frac{| \left\langle P^T C_h(\Phi \boldsymbol{a}), (M^T P)^{-1} M^T \Phi \boldsymbol{a} \right\rangle |}{||(M^T P)^{-1} M^T \Phi \boldsymbol{a}||} \cdot ||(P^T M)^{-1}||.$$

Since $\mathcal{M}_\mathcal{F} \subset \mathcal{M}$ is a linear subspace of $\mathcal{M}$ of dimension $m - 1$, we can write an orthogonal basis for $\mathcal{M}_\mathcal{F}$ as $M K_M \in \mathbb{R}^{N \times m-1}$ for some $K_M \in \mathbb{R}^{m \times m-1}$, where every basis vector of $\mathcal{M}_\mathcal{F}$ is a linear combination of the basis vectors in M. Since $\langle \boldsymbol{u}, \Phi \boldsymbol{a} \rangle = 0 \ \forall \boldsymbol{u} \in \mathcal{M}_\mathcal{F}$ we can state:

$$\left\langle P^T C_h(\Phi \boldsymbol{a}), (M^T P)^{-1} M^T \Phi \boldsymbol{a} \right\rangle = \left\langle M (P^T M)^{-1} P^T C_h(\Phi \boldsymbol{a}), \Phi \boldsymbol{a} \right\rangle$$
$$= \left\langle M (P^T M)^{-1} (P^T C_h(\Phi \boldsymbol{a}) - (P^T M) K_M K_M^T M^T C_h(\Phi \boldsymbol{a})), \Phi \boldsymbol{a} \right\rangle$$
$$= \left\langle M (P^T M)^{-1} P^T \left(I - M K_M K_M^T M^T \right) C_h(\Phi \boldsymbol{a}), \Phi \boldsymbol{a} \right\rangle$$
$$= \left\langle P^T \left(I - M K_M K_M^T M^T \right) C_h(\Phi \boldsymbol{a}), (M^T P)^{-1} M^T \Phi \boldsymbol{a} \right\rangle.$$

Clearly, $M K_M K_M^T M^T C_h(\Phi \boldsymbol{a})$ is the orthogonal projection of $C_h(\Phi \boldsymbol{a})$ on $\mathcal{M}_\mathcal{F}$. Using the Cauchy-Schwartz inequality we find:

$$||\varepsilon_{\mathrm{CLS}}|| \leq ||\varepsilon_{\mathrm{DEIM}}|| + ||P^T \left(I - M K_M K_M^T M^T \right) C_h(\Phi \boldsymbol{a})|| \cdot ||(P^T M)^{-1}||.$$

Finally, substituting the bound for $||\varepsilon_{\mathrm{DEIM}}||$ found in [24] and using that $||P^T|| = 1$ we obtain the desired inequality.

Appendix D. Fully discrete energy-conservation with energy-conserving DEIM and Runge-Kutta methods

In this appendix we prove (29). The analysis of a change in the norm of a numerical solution from one time step to the next using Runge-Kutta (RK) methods is a classical analysis [40,51]. In its generality a Runge-Kutta method for a hROM takes the form:

$$\boldsymbol{A}_i = \boldsymbol{a}^n + \Delta t \sum_{j=1}^{s} \alpha_{ij} \left(-\Phi^T M \boldsymbol{c}(\boldsymbol{A}_j) + \nu D_r \boldsymbol{A}_j \right) \tag{D.1}$$

$$\boldsymbol{a}^{n+1} = \boldsymbol{a}^n + \Delta t \sum_{i=1}^{s} \beta_i \left(-\Phi^T M \boldsymbol{c}(\boldsymbol{A}_i) + \nu D_r \boldsymbol{A}_i \right). \tag{D.2}$$

Here, $\alpha_{ij}, \beta_i \in \mathbb{R}$ and $s \in \mathbb{N}$ are coefficients characterizing the specific RK method. We denote $f_i := -\Phi^T M \boldsymbol{c}(\boldsymbol{A}_i) + \nu D_r \boldsymbol{A}_i$. An expression for the change in norm $\left\| \boldsymbol{a}^{n+1} \right\|^2 - ||\boldsymbol{a}^n||^2$ can be found in [40, theorem IV-2.2] and is given as:

$$\left\| \boldsymbol{a}^{n+1} \right\|^2 - ||\boldsymbol{a}^n||^2 = 2\Delta t \sum_{i=1}^{s} \beta_i \langle \boldsymbol{A}_i, f_i \rangle - \Delta t^2 \sum_{i,j=1}^{s} m_{ij} \langle f_i, f_j \rangle, \tag{D.3}$$

where:

$$m_{ij} = \beta_i \alpha_{ij} + \beta_j \alpha_{ji} - \beta_i \beta_j, \quad i,j \in \{1, ..., s\}.$$

The reduced total kinetic energy then changes as:

$$K_r^{n+1} - K_r^n = \Delta t \sum_{i=1}^{s} \beta_i \langle \boldsymbol{A}_i, f_i \rangle - \frac{\Delta t^2}{2} \sum_{i,j=1}^{s} m_{ij} \langle f_i, f_j \rangle. \tag{D.4}$$

Classical energy-conserving RK methods (like in [87,86]) satisfy $m_{ij} = 0 \; \forall i, j \in \{1, ..., s\}$ so that the second term on the right hand side equals zero. Explicit energy-conserving RK methods similarly assure that this term equals zero [51,79]. The first term in (D.4) can be written as:

$$\Delta t \sum_{i=1}^{s} \beta_i \langle \boldsymbol{A}_i, f_i \rangle = -\Delta t \sum_{i=1}^{s} \beta_i \langle \boldsymbol{A}_i, \Phi^T M \boldsymbol{c}(\boldsymbol{A}_i) \rangle + \Delta t \sum_{i=1}^{s} \beta_i \nu \langle \boldsymbol{A}_i, D_r \boldsymbol{A}_i \rangle.$$

Using an energy-conserving DEIM algorithm the first term on the right-hand side will be zero as condition (24) is satisfied for these methods. The second term on the right-hand side can be rewritten using the properties of the reduced diffusion operator to obtain:

$$\Delta t \sum_{i=1}^{s} \beta_i \langle \boldsymbol{A}_i, f_i \rangle = -\Delta t \sum_{i=1}^{s} \beta_i \nu \left\| Q_r \boldsymbol{A}_i \right\|^2,$$

completing the proof.

References

[1] B.M. Afkham, J.S. Hesthaven, Structure-preserving model-reduction of dissipative Hamiltonian systems, J. Sci. Comput. 81 (1) (2019) 3–21.
[2] B.M. Afkham, N. Ripamonti, Q. Wang, J.S. Hesthaven, Conservative Model Order Reduction for Fluid Flow, Springer International Publishing, Cham, 2020, pp. 67–99.
[3] S.E. Ahmed, O. San, Breaking the Kolmogorov barrier in model reduction of fluid flows, Fluids 5 (1) (2020).
[4] N. Akkari, F. Casenave, V. Moureau, Time stable reduced order modeling by an enhanced reduced order basis of the turbulent and incompressible 3d Navier–Stokes equations, Math. Comput. Appl. 24 (2019) 45.
[5] D. Amsallem, C. Farhat, Stabilization of projection-based reduced-order models, Int. J. Numer. Methods Eng. 91 (4) (2012) 358–377.
[6] V.I. Arnol'd, The Hamiltonian nature of the Euler equations in the dynamics of a rigid body and of an ideal fluid, Usp. Mat. Nauk 24 (3(147)) (1969) 225–226.
[7] P. Astrid, S. Weiland, K. Willcox, T. Backx, Missing point estimation in models described by proper orthogonal decomposition, IEEE Trans. Autom. Control 53 (10) (2008) 2237–2251.
[8] M. Balajewicz, E.H. Dowell, Stabilization of projection-based reduced order models of the Navier-Stokes, Nonlinear Dyn. 70 (2) (2012) 1619–1632.
[9] M. Balajewicz, E.H. Dowell, B. Noack, Low-dimensional modelling of high Reynolds number shear flows incorporating constraints from the Navier-Stokes equation, J. Fluid Mech. 729 (2013) 285.
[10] J. Barnett, C. Farhat, Quadratic approximation manifold for mitigating the Kolmogorov barrier in nonlinear projection-based model order reduction, J. Comput. Phys. 464 (2022) 111348.
[11] M. Barrault, Y. Maday, N.C. Nguyen, A.T. Patera, An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, C. R. Math. 339 (9) (2004) 667–672.
[12] P. Benner, P. Goyal, J. Heiland, I. Pontes Duff, Operator inference and physics-informed learning of low-dimensional models for incompressible flows, arXiv: 2010.06701 [math.DS], 2020.
[13] P. Blonigan, K. Carlberg, F. Rizzi, M. Howard, J. Fike, Model reduction for hypersonic aerodynamics via conservative LSPG projection and hyper-reduction, AIAA Paper 2020-0104, AIAA Scitech 2020 Forum, Orlando, FL, 6-10 January 2020.
[14] J. Borggaard, A. Hay, D. Pelletier, Interval-based reduced-order models for unsteady fluid flow, Int. J. Numer. Anal. Model. 4 (2007) 353–367.
[15] S. Boyd, S.P. Boyd, L. Vandenberghe, Convex Optimization, Berichte über Verteilte Messysteme, vol. 1, Cambridge University Press, 2004.
[16] S.L. Brunton, J.N. Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control, Cambridge University Press, 2022.
[17] K. Carlberg, M. Barone, H. Antil, Galerkin v. least-squares Petrov—Galerkin projection in nonlinear model reduction, J. Comput. Phys. 330 (2017) 693–734.

[18] K. Carlberg, C. Bou-Mosleh, C. Farhat, Efficient non-linear model reduction via a least-squares Petrov-Galerkin projection and compressive tensor approxima-tions, Int. J. Numer. Methods Eng. 86 (2011) 155.

[19] K. Carlberg, Y. Choi, S. Sargsyan, Conservative model reduction for finite-volume models, J. Comput. Phys. 371 (2017) 11.

[20] K. Carlberg, C. Farhat, J. Cortial, D. Amsallem, The gnat method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows, J. Comput. Phys. 242 (2013) 623–647.

[21] J. Chan, Entropy stable reduced order modeling of nonlinear conservation laws, J. Comput. Phys. 423 (dec 2020) 109789.

[22] S. Chaturantabut, Temporal localized nonlinear model reduction with a priori error estimate, Appl. Numer. Math. 119 (2017) 225–238.

[23] S. Chaturantabut, C. Beattie, S. Gugercin, Structure-preserving model reduction for nonlinear port-Hamiltonian systems, SIAM J. Sci. Comput. 38 (5) (2016) B837–B865.

[24] S. Chaturantabut, D.C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. Sci. Comput. 32 (2010) 2737–2764.

[25] S. Chaturantabut, D.C. Sorensen, A state space error estimate for POD-DEIM nonlinear model reduction, SIAM J. Numer. Anal. 50 (2012) 46–63.

[26] Saifon Chaturantabut, Nonlinear Model Reduction via Discrete Empirical Interpolation, PhD thesis, Rice University, 2011.

[27] D.M. Copeland, S.W. Cheung, K. Huynh, Y. Choi, Reduced order models for Lagrangian hydrodynamics, Comput. Methods Appl. Mech. Eng. 388 (2022) 114259.

[28] P.A. Davidson, Turbulence: An Introduction for Scientists and Engineers, OUP, Oxford, 2004.

[29] M. D'Elia, M. Gunzburger, G. Rozza, Quantification of Uncertainty: Improving Efficiency and Technology, QUIET Selected Contributions, 2020.

[30] R. Everson, L. Sirovich, Karhunen–Loève procedure for gappy data, J. Opt. Soc. Am. A 12 (8) (Aug 1995) 1657–1664.

[31] C. Farhat, P. Avery, T. Chapman, J. Cortial, Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency, Int. J. Numer. Methods Eng. 98 (2014) 06.

[32] C. Farhat, T. Chapman, 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, Int. J. Numer. Methods Eng. 102 (2015) 1077–1110.

[33] L. Fick, Y. Maday, A.T. Patera, T. Taddei, A stabilized POD model for turbulent flows over a range of Reynolds numbers: optimal parameter sampling and constrained projection, J. Comput. Phys. 371 (2018) 214–243.

[34] R. Geelen, S. Wright, K. Willcox, Operator inference for non-intrusive model reduction with quadratic manifolds, Comput. Methods Appl. Mech. Eng. 403 (2023) 115717.

[35] G. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215.

[36] Y. Gong, Q. Wang, Z. Wang, Structure-preserving Galerkin POD reduced-order modeling of Hamiltonian systems, Comput. Methods Appl. Mech. Eng. 315 (2017) 780–798.

[37] M.A. Grepl, Y. Maday, N.C. Nguyen, A.T. Patera, Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations, ESAIM: M2AN 41 (3) (2007) 575–605.

[38] S. Grimberg, C. Farhat, R. Tezaur, C. Bou-Mosleh, Mesh sampling and weighting for the hyperreduction of nonlinear Petrov–Galerkin reduced-order models with local reduced-order bases, Int. J. Numer. Methods Eng. 122 (2020) 12.

[39] S. Grimberg, C. Farhat, N. Youkilis, On the stability of projection-based model order reduction for convection-dominated laminar and turbulent flows, J. Comput. Phys. 419 (2020) 109681.

[40] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations, Springer Series in Computational Mathematics, Springer, 2002.

[41] M. Hanke, Limitations of the L-curve method in ill-posed problems, BIT Numer. Math. 36 (1996) 287–301.

[42] P.C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, SIAM Rev. 34 (4) (1992) 561–580.

[43] D. Hartmann, M. Herz, U. Wever, Model Order Reduction a Key Technology for Digital Twins, 2018, pp. 167–179.

[44] T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics, Springer, 2001.

[45] J.S. Hesthaven, C. Pagliantini, Structure-preserving reduced basis methods for Poisson systems, Math. Comput. 90 (330) (July 2021) 1701–1740.

[46] J.S. Hesthaven, C. Pagliantini, N. Ripamonti, Structure-preserving model order reduction of Hamiltonian systems, arXiv:2109.12367 [math.NA], 2021.

[47] P. Holmes, J.L. Lumley, G. Berkooz, C.W. Rowley, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge Monographs on Mechanics, Cambridge University Press, 2012.

[48] C. Huang, C. Wentland, K. Duraisamy, C. Merkle, Model reduction for multi-scale transport problems using structure-preserving least-squares projections with variable transformation, arXiv:2011.02072 [physics.comp-ph], 2020.

[49] W.L. IJzerman, Signal Representation and Modeling of Spatial Structures in Fluids, PhD thesis, University of Twente, April 2000.

[50] I. Kalashnikova, S. Arunajatesan, A stable Galerkin reduced order model for compressible flow, in: Blucher Mechanical Engineering Proceedings, vol. 1, 2014, pp. 1399–1423.

[51] D.I. Ketcheson, Relaxation Runge-Kutta methods: conservation and stability for inner-product norms, SIAM J. Numer. Anal. 57 (6) (2019) 2850–2870.

[52] R. Knikker, Study of a staggered fourth-order compact scheme for unsteady incompressible viscous flows, Int. J. Numer. Methods Fluids 59 (2009).

[53] D.A. Knoll, D.E. Keyes, Jacobian-free Newton–Krylov methods: a survey of approaches and applications, J. Comput. Phys. 193 (2) (2004) 357–397.

[54] E. Kreyszig, Introductory Functional Analysis with Applications, Wiley Classics Library, Wiley, 1991.

[55] T. Lassila, A. Manzoni, A. Quarteroni, G. Rozza, Model Order Reduction in Fluid Dynamics: Challenges and Perspectives, Springer International Publishing, Cham, 2014, pp. 235–273.

[56] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Society for Industrial and Applied Mathematics, 1995.

[57] K. Lee, K.T. Carlberg, Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders, J. Comput. Phys. 404 (2020) 108973.

[58] D.H. Li, M. Fukushima, A derivative-free line search and global convergence of Broyden-like method for nonlinear equations, Optim. Methods Softw. 13 (3) (2000) 181–201.

[59] J. Loiseau, S.L. Brunton, Constrained sparse Galerkin regression, J. Fluid Mech. 838 (2018) 42–67.

[60] T. Lu, S. Shiou, Inverses of 2×2 block matrices, Comput. Math. Appl. 43 (1) (2002) 119–129.

[61] K. Manohar, B.W. Brunton, J.N. Kutz, S.L. Brunton, Data-driven sparse sensor placement for reconstruction: demonstrating the benefits of exploiting known patterns, IEEE Control Syst. Mag. 38 (3) (2018) 63–86.

[62] R. Maulik, B. Lusch, P. Balaprakash, Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders, Phys. Fluids 33 (3) (Mar 2021) 037106.

[63] S.A. McQuarrie, S. Huang, K.E. Willcox, Data-driven reduced-order models via regularised operator inference for a single-injector combustion process, J. R. Soc. N.Z. 51 (2) (2021) 194–211.

[64] M.F. Barone, I. Kalashnikova, D.J. Segalman, H.K. Thornquist, Stable Galerkin reduced order models for linearized compressible flow, J. Comput. Phys. 228 (6) (2009) 1932–1946.

[65] M.L. Minion, D.L. Brown, Performance of under-resolved two-dimensional incompressible flow simulations, II, J. Comput. Phys. 138 (2) (1997) 734–765.

[66] Y. Miyatake, Structure-preserving model reduction for dynamical systems with a first integral, Jpn. J. Ind. Appl. Math. 36 (3) (2019) 1021–1037.

[67] M. Mohebujjaman, L. Rebholz, T. Iliescu, Physically-constrained data-driven correction for reduced order modeling of fluid flows, Int. J. Numer. Methods Fluids 89 (2018) 09.

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

[69] N.T. Mücke, S.M. Bohté, C.W. Oosterlee, Reduced order modeling for parameterized time-dependent PDEs using spatially and memory aware deep learning, J. Comput. Sci. 53 (2021) 101408.

[70] A. Nishida, Experience in Developing an Open Source Scalable Software Infrastructure in Japan, vol. 6017, 2010, pp. 448–462.

[71] B.R. Noack, M. Morzynski, G. Tadmor, Reduced-Order Modelling for Flow Control, CISM International Centre for Mechanical Sciences, Springer, Vienna, 2011.

[72] P.J. Olver, A nonlinear Hamiltonian structure for the Euler equations, J. Math. Anal. Appl. 89 (1) (1982) 233–250.

[73] C. Pagliantini, F. Vismara, Gradient-preserving hyper-reduction of nonlinear dynamical systems via discrete empirical interpolation, SIAM J. Sci. Comput. 45 (5) (2023).

[74] B. Peherstorfer, D. Butnaru, K. Willcox, H. Bungartz, Localized discrete empirical interpolation method, SIAM J. Sci. Comput. 36 (1) (2014) A168–A192.

[75] B. Peherstorfer, Z. Drmač, S. Gugercin, Stability of discrete empirical interpolation and gappy proper orthogonal decomposition with randomized and deterministic sampling points, SIAM J. Sci. Comput. 42 (5) (2020) A2837–A2864.

[76] B. Peherstorfer, K. Willcox, Data-driven operator inference for nonintrusive projection-based model reduction, Comput. Methods Appl. Mech. Eng. 306 (2016) 196–215.

[77] L. Peng, K. Mohseni, Symplectic model reduction of Hamiltonian systems, SIAM J. Sci. Comput. 38 (1) (2016) A1–A27.

[78] R. Polyuga, A. Schaft, Effort and flow-constraint reduction methods for structure preserving model reduction of port-Hamiltonian systems, Syst. Control Lett. 61 (2012) 03.

[79] H. Ranocha, M. Sayyari, L. Dalcin, M. Parsani, D.I. Ketcheson, Relaxation Runge-Kutta methods: fully discrete explicit entropy-stable schemes for the compressible Euler and Navier-Stokes equations, SIAM J. Sci. Comput. 42 (2) (2020) A612–A638.

[80] F. Romor, G. Stabile, G. Rozza, Non-linear manifold rom with convolutional autoencoders and reduced over-collocation method, arXiv:2203.00360 [math.NA], 2022.

[81] C.W. Rowley, T. Colonius, R.M. Murray, Model reduction for compressible flows using POD and Galerkin projection, Phys. D: Nonlinear Phenom. 189 (1) (2004) 115–129.

[82] G. Rozza, M. Hess, G. Stabile, M. Tezzele, F. Ballarin, Basic Ideas and Tools for Projection-Based Model Reduction of Parametric Partial Differential Equations, De Gruyter, 2020, pp. 1–47.

[83] A.K. Saibaba, Randomized discrete empirical interpolation method for nonlinear model reduction, SIAM J. Sci. Comput. 42 (3) (2020) A1582–A1608.

[84] O. San, A. Staples, High-order methods for decaying two-dimensional homogeneous isotropic turbulence, Comput. Fluids 63 (2012) 12.

[85] B. Sanderse, Energy-Conserving Discretization Methods for the Incompressible Navier-Stokes Equations: Application to the Simulation of Wind-Turbine Wakes, PhD thesis, Centrum voor Wiskunde en Informatica, 2013.

[86] B. Sanderse, Energy-conserving Runge–Kutta methods for the incompressible Navier–Stokes equations, J. Comput. Phys. 233 (2013) 100–131.

[87] B. Sanderse, Non-linearly stable reduced-order models for incompressible flow with energy-conserving finite volume methods, J. Comput. Phys. 421 (2020) 109736.

[88] B. Sanderse, B. Koren, Accuracy analysis of explicit Runge–Kutta methods applied to the incompressible Navier–Stokes equations, J. Comput. Phys. 231 (8) (2012) 3041–3063.

[89] C. Sanderson, R. Curtin, Armadillo: a template-based C++ library for linear algebra, J. Open Sour. Softw. 1 (2016) 26.

[90] C. Sanderson, R. Curtin, A User-Friendly Hybrid Sparse Matrix Class in C++, 2018, pp. 422–430.

[91] A.M. Schein, K.T. Carlberg, M.J. Zahr, Preserving general physical properties in model reduction of dynamical systems via constrained-optimization projection, arXiv:2011.13998 [abs], 2020.

[92] H. Sharma, Z. Wang, B. Kramer, Hamiltonian operator inference: physics-preserving learning of reduced-order models for canonical Hamiltonian systems, Phys. D: Nonlinear Phenom. 431 (2022) 133122.

[93] L. Sirovich, Turbulence and the dynamics of coherent structures. I - coherent structures. II - symmetries and transformations. III - dynamics and scaling, Q. Appl. Math. 45 (1987) 10.

[94] B.E. Sonday, A. Singer, C.W. Gear, I.G. Kevrekidis, Manifold learning techniques and model reduction applied to dissipative PDEs, arXiv:1011.5197 [physics.comp-ph], 2010.

[95] I. Tezaur, M. Barone, Stable and efficient Galerkin reduced order models for non-linear fluid flow, in: 6th AIAA Theoretical Fluid Mechanics Conference, 2011.

[96] A. Treuille, A. Lewis, Z. Popovic, Model reduction for real-time fluids, ACM Trans. Graph. 25 (2006) 826.

[97] Z. Wang, Structure-preserving Galerkin POD-DEIM reduced-order modeling of Hamiltonian systems, arXiv:2103.00388 [abs], 2021.

[98] P. Wesseling, Principles of Computational Fluid Dynamics, Springer Series in Computational Mathematics, Springer, 2001.

[99] X. Xie, M. Mohebujjaman, L. Rebholz, T. Iliescu, Data-driven filtered reduced order modeling of fluid flows, SIAM J. Sci. Comput. 40 (2017) 09.

[100] R. Zimmermann, K. Willcox, An accelerated greedy missing point estimation procedure, SIAM J. Sci. Comput. 38 (5) (2016) A2827–A2850.