Geometric Model Order Reduction

THIS IS A TEMPORARY TITLE PAGE
It will be replaced for the final print by a version
provided by the service academique.

Thèse n. XXXX XXXX
présenté le 10 septembre 2018
à la Faculté des Sciences de Base
laboratoire MCSS
École Polytechnique Fédérale de Lausanne

pour l'obtention du grade de Docteur ès Sciences
par

Babak Maboudi Afkham

acceptée sur proposition du jury:

Prof. Marc Troyanov, président du jury
Prof. Jan S. Hesthaven, directeur de thèse
Prof. Fabio Nobile, rapporteur
Prof. Mario Ohlberger, rapporteur
Dr. Kevin Carlberg, rapporteur

Lausanne, EPFL, 2018

Abstract

During the past decade, model order reduction (MOR) has been successfully applied to reduce the computational complexity of elliptic and parabolic systems of partial differential equations (PDEs). However, MOR of hyperbolic equations remains a challenge. Symmetries and conservation laws, which are a distinctive feature of such systems, are often destroyed by conventional MOR techniques, resulting in a perturbed and often unstable reduced system. The goal of this thesis is to study and develop model order reduction techniques that can preserve nonlinear invariants, symmetries, and conservation laws and to understand the stability properties of these methods compared to conventional techniques. Hamiltonian systems, as systems that are driven by symmetries, are studied intensively from the point of view of MOR. Furthermore, a conservative model reduction of fluid flow is presented. It is illustrated that conserving invariants, conservation laws, and symmetries not only result in a physically meaningful reduced system, but also result in an accurate and robust reduced system with enhanced stability.

Keywords: Model Order Reduction, Structure-Preserving, The Greedy Basis Generation, Symplectic Galerkin Method, Weighted Norm, Hamiltonian Systems, Skew-Symmetric Formulation, Symplectic Geometry.

Résumé

Au cours de la dernière décennie, la réduction d'ordre de modèle (ROM) a réussi à réduire la complexité de calcul des systèmes elliptiques et paraboliques d'équations aux dérivées partielles (EDP). Cependant, ROM des équations hyperboliques reste un défi. Les symétries et les lois de conservation, qui sont une caractéristique distinctive de tels systèmes, sont souvent détruites par les techniques conventionnelles de ROM qui aboutissent à un système réduit perturbé et souvent instable. Le but de cette thèse est d'étudier et de développer des techniques de réduction d'ordre de modèle pouvant préserver les invariants nonlinéaires, les symétries et les lois de conservation et de comprendre les propriétés de stabilité de ces méthodes par rapport aux techniques conventionnelles. Les systèmes Hamiltoniens, en tant que systèmes pilotés par des symétries, sont étudiés de maniére intensive depuis le point de vue ROM. De plus, une réduction modérée du débit de fluide est présentée. Il est illustré que la conservation des invariants, des lois de conservation et des symétries non seulement aboutit à un système réduit physiquement significatif, mais construit également un système réduit robuste avec une stabilité accrue.

Mots-clés: Réduction d'Ordre de Modèle, Conservation de Structure, Génération de Bases, Méthode de Galerkin Symplectique, Norme Pondérée, Systèmes Hamiltoniens, Formulation Asymétrique, Géométrie Symplectique.

Contents

1 Introduction

Mathematical modeling and scientific computing has become an inseparable part of engineering and science, thanks to advances in computational science and technology. Models expressed as partial differential equations (PDEs) can be found in a wide range of disciplines from social sciences, biology, cosmology, modern and classical physics, and engineering to industrial applications. The overwhelming success of such models for approximately describing nature has encouraged the development of complex mathematical models in order to attain higher accuracy in representing physical phenomena. The complexity of many modern applications, however, is computationally prohibitive with classical approaches. The curse of dimensionality for multi-dimensional parameter sets, i.e. the exponential growth in the computational costs for problems in higher dimensions, is an example of a computational inefficiency that inhibits progress.

Reduced order modeling (ROM), as opposed to high-fidelity modeling, has emerged as a successful attempt to reduce the intrinsic computational inefficiencies of modern complex models [52, 86, 14, 4]. ROM aims to accurately represent a high-dimensional model with a few degrees of freedom, by exploiting empirical or physical structure in data. As a result of confining the model to only these degrees of freedom, the computational costs can be substantially reduced. Although these methods do not eliminate the need for high-fidelity modeling, they significantly accelerate the evaluation of outputs of interest when the repeated evaluation of the high-fidelity model is required [52].

The recognition of patterns in data makes ROM comparable with machine learning techniques, e.g. in computer science and statistics developed during the past decades [74]. However, the deterministic nature of PDEs, Combined with the control in the choice of data generation process, gives ROM a distinct take.

This difference between ROM and conventional machine learning techniques becomes more apparent with time-dependent problems. Time-symmetries of high-fidelity models are lost in the assembly of data, which sometimes, result in an ill-represented ROM

[4]. Although this inaccuracy in the representation is less evident for parabolic PDEs, the development of ROMs for hyperbolic PDEs, where symmetries are a fundamental feature, remains a challenge [2, 60, 36, 10, 24, 9, 81].

The main aim of this thesis is to develop ROM techniques that capture symmetries in a system of PDEs. The conservation of such symmetries not only results in a robust ROM, but helps with the construction of a meaningful reduced order model. Time, as a parameter, plays a crucial role in the existence and the conservation of symmetries. Therefore, this thesis puts a particular emphasis on the treatment of the time variable by studying systems that depend on none, or otherwise very small number of additional parameters. This might, at first glance, sounds counter intuitive in the context ROM. However, the isolation of time provides remarkable insight into the theory of ROM, and into the mathematical modeling. Nevertheless, the main results of this thesis can be extended to the general parameter setting while retaining all the benefits associated with the conservation of symmetries.

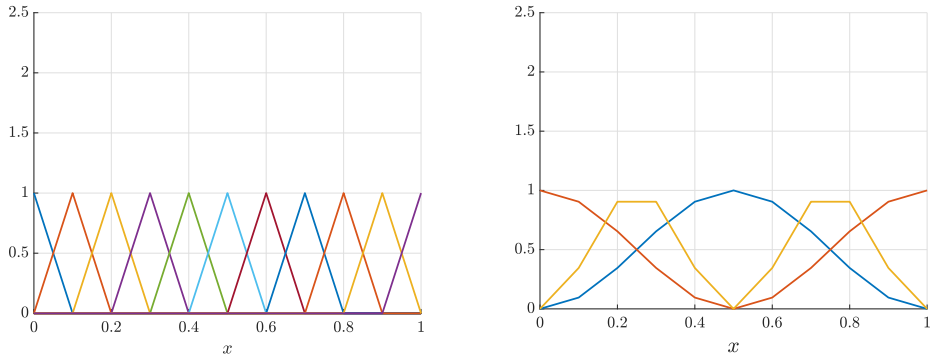
In what is left of this chapter, we discuss the difficulties involved in treating time as a parameter in the context of ROM, and conclude by briefly discussing the content of the thesis.

1.1 Space and Time in ROM

Reduced basis (RB) methods are among the most popular techniques to develop ROMs. These methods have been successful in reducing the computational complexity of large scale systems of partial differential equations, and have been used in many disciplines in engineering and science and also applied in industry. Many applications of RB methods can be found in [52, 86, 14, 4] and the references therein.

RB methods are based on the assumption that a state of a solution to a system of PDEs can be well approximated by a few degrees of freedom, chosen from a low-dimensional linear subspace. A basis for this subspace, referred to as the *reduced basis*, is constructed to accurately represent the state of the system for a particular empirical setting. A projection operator, often of a linear type, is constructed to confine the state of the system to this subspace. This constructs a new system of partial differential equations that is described by a few independent variables. In principle, this system can be evaluated at an accelerated rate, compared to the high-fidelity system.

In the context of the finite element method (FEM) where a solution to a PDE is described as a linear combination of basis functions, RB methods can be represented visually. Consider the equation governing a one-dimensional wave in a periodic do-



(a) FEM basis functions

(b) RB basis functions

Figure 1.1 – spatial representation of MOR

main.

$$\frac{\partial^2}{\partial t^2}u(t,x) + \frac{\partial^2}{\partial x^2}u(t,x) = 0, \quad x \in [0,1], \tag{1.1}$$

together with some initial condition $u(t,x) = u_0(x)$. A standard finite element discretization requires u to be a linear combination of n, time and problem independent, basis functions $\varphi_i(x)$ as

$$u(x,t) \approx \sum_{i=1}^{n} c_i(t)\varphi_i(x), \tag{1.2}$$

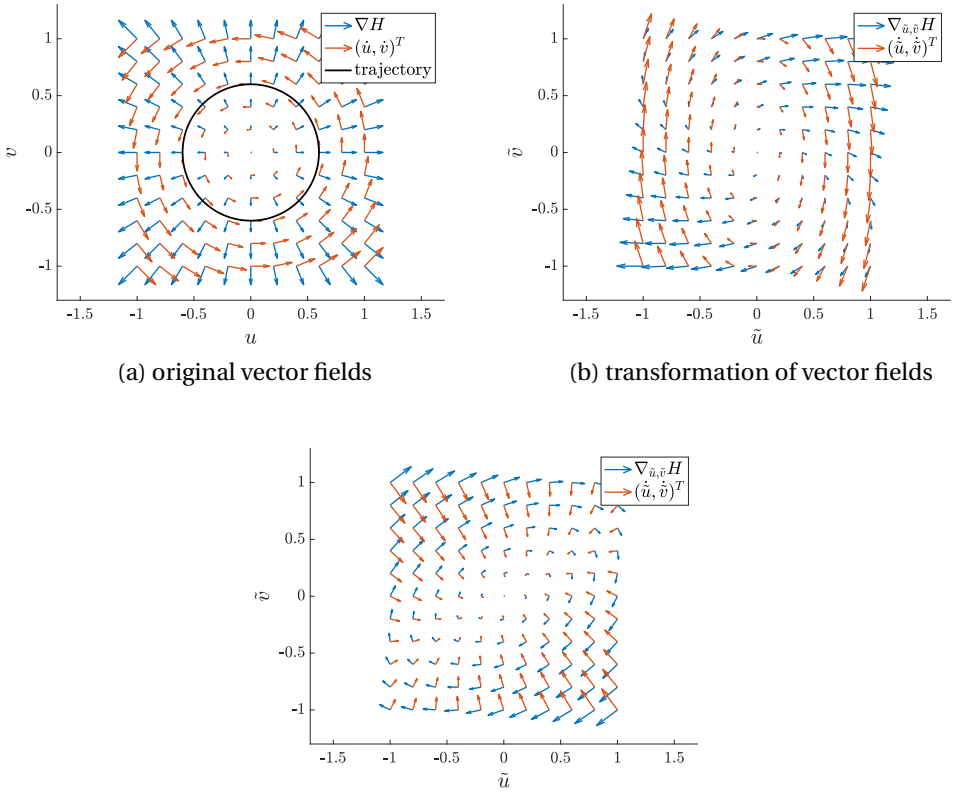
where $c_i(t)$ are the expansion coefficients that yield the degrees of freedom. In this setting, the choice of φ_i is independent of (1.1). Exploiting some problem-specific patterns, e.g., in the initial or the boundary conditions, problem geometry, or numerical properties, allows us to introduce a new, but potentially smaller, set of basis functions ψ_i. By

$$u(x,t) \approx \sum_{i=1}^{k} c_i'(t)\psi_i(x), \tag{1.3}$$

where ψ_i are chosen such that (1.3) delivers comparable accuracy to (1.2). If $k \ll n$, we gain acceleration by evaluating only k coefficients rather than n. Often, RB methods require the relation between ψ_i and φ_i to be a linear relation, i.e.

$$\psi_i = \sum_{j=1}^{n} r_{ij}\varphi_j, \quad i = 1,\dots,k. \tag{1.4}$$

A rough sketch of basis functions φ_i and ψ_i is illustrated in Figure 1.1. This is a *spatial perspective* of reduced order modeling.



(a) original vector fields

(b) transformation of vector fields

(c) symplectic transformation of vector fields

Figure 1.2 – temporal representation of MOR

To be able to visualize the temporal aspect of MOR, we simplify (1.1) to obtain an ordinary differential equation (ODE) $\ddot{u}(t) - u(t) = 0$, which can be expressed in terms of first order ODEs as

$$\begin{cases} \dot{u}(t) = v(t), \\ \dot{v}(t) = -u(t). \end{cases} \tag{1.5}$$

Introduction of the new variable v is not only an algebraic tool, but carries a deeper insight into the dynamics of the original second order ODE. For instance, $H(u,v) = u^2 + v^2$ is a constant quantity, and can be interpreted as the energy of (1.5). Therefore, the value of v is restricted by H in a nonlinear sense. Another insight into the dynamics of (1.5) is revealed when we consider vector fields in (u,v) coordinate system, commonly referred to as the *phase space*. Figure 1.2a shows the vectors fields ∇H and $(\dot{u},\dot{v})^T$. We immediately notice the orthogonality of the two vector fields. Periodic behaviour of (1.5) is a result of this delicate relation. Such properties of (1.5) that are unchanged along a trajectories of (1.5) are often referred to as *symmetries* of (1.5).

Let us now study the symmetries of (1.5) in a transformed coordinate system. Fig-

ure 1.2b shows the transformation of ∇H and $(\dot{u}, \dot{v})^T$ over some linear transformation $(\tilde{u}, \tilde{v}) = T(u, v)$. We notice that the orthogonality of the two vector fields is destroyed. Although, the dynamics of the original and the transformed system are the same, the transformed system carries less symmetry. For some class of linear transformations, *symplectic transformations*, the orthogonality of the two vector fields is preserved. This can be seen in Figure 1.2c where a linear symplectic transformation is applied to ∇H and $(\dot{u}, \dot{v})^T$.

In a numerical approximation, loss of symmetries can have profound consequences for the overall dynamics of a system. For example, the periodic trajectory of (1.5), which is a distinctive feature of the dynamics, may no longer remain periodic in a non-symmetric coordinate system.

In the context of MOR basis functions, e.g., those in Figure 1.1a, can be viewed as a basis for the phase space. Subsequently, a solution expanded in this basis can be translated as a vector in the phase space. The relation (1.4) is therefore interpreted as a change in the coordinate system. This is the *temporal perspective* of MOR.

Similar to (1.5), we can define orthogonal vectors fields for (1.1) as ∇H and $(\partial u/\partial t, \partial v/\partial t)^T$ with $v = \partial u/\partial t$ and $H(u, v) = \int v^2 + (\partial u/\partial x)^2 \, dx$. Therefore, we expect a general RB method to result in a non-symmetric phase space. In particular since the patterns in the ensemble of solutions to (1.1) does not reveal the subtle relation between the two vector fields.

Nonlinear invariants and symmetries, such as those discussed above, are a fundamental feature of hyperbolic system of PDEs. Loss of symmetries in such systems can help explain challenges in MOR of hyperbolic problems.

The main aim of this thesis is to seek RB techniques that construct a reduced phase space that captures the symmetries of the high-fidelity system of PDEs. This ensures conservation of some nonlinear invariants and, subsequently, a good approximation of the overall dynamics in the reduced system. Hamiltonian systems, as systems that are driven by symmetries, are studied intensively from MOR view point. We then develop methods that preserve symmetries of a fluid flow.

1.2 Overview of The Thesis

The overall goal of this thesis is to study and develop RB techniques that preserve nonlinear invariants, symmetries, and conservation laws. Furthermore, it aims to understand the stability and robustness properties of these methods, compared to conventional RB techniques. This thesis mainly focuses on the MOR of time-dependent and, in particular, hyperbolic PDEs. A particular emphasis is put on model order reduction of Hamiltonian systems to understand the role of time in structure-preserving

MOR. The main results are then generalized to construct MOR methods for fluid flow.

Chapter 2 surveys the background on smooth manifolds and Hamiltonian systems. We introduce the concept of geometric symmetry and how it relates to the dynamics of a time-dependent differential equations. We also briefly introduce methods for conserving these symmetries in a numerical evaluation.

An overview of the theory of model order reduction is presented in Chapter 3. We present conventional RB techniques, e.g. proper orthogonal decomposition and the greedy method, for generating an accurate reduced basis. Galerkin and Petrov-Galerkin projection is then discussed to construct a reduced system. This chapter discusses the efficient evaluation of nonlinear terms by introducing the empirical interpolation method.

Symplectic MOR for Hamiltonian systems is developed in Chapter 4. It is discussed how symplectic transformations can be used to construct a reduced Hamiltonian system that preserves the dynamics of the high-fidelity Hamiltonian system. We present a greedy method for the generation of a symplectic basis as well as other SVD-based symplectic model reduction techniques. Accuracy, stability, and efficiency of the method are discussed and illustrated by numerical experiments.

In Chapter 5 we couple the symplectic model order reduction with a weighted inner-product. We show that this can be viewed as a natural generalization of the symplectic MOR. Numerical experiments are presented to illustrate how this method can be beneficial when an unstructured numerical discretization is used in the high-fidelity system.

Chapter 6 presents symplectic MOR in the context of dissipative Hamiltonian system. It is discussed how a canonical extension of the dissipative Hamiltonian system yields a closed and conservative system. An application of a symplectic MOR on an extended system can help with a correct evolution of the Hamiltonian. The performance of this method is illustrated through simulations of dissipative Hamiltonian and port-Hamiltonian systems.

A conservative model reduction of fluid flow is presented Chapter 7. It is explained how the skew-symmetry of differential operators in a fluid flow can help to recover the conservation of quadratic invariants, e.g., energy, at the level of reduced system. It is discussed how this gives rise to a physically meaningful reduced system with quadratic invariants with respect to the reduced variables. Stability properties of the method is illustrated through various numerical experiments of incompressible and compressible fluids.

2 Symplectic Geometry and Hamiltonian systems

In preparation for later chapters, we recall basic background regarding differential geometry and symplectic geometry. Although the applications of MOR, as presented in this thesis are mostly described on a linear vector space, the theory of smooth manifold can provide a substantial insight into the geometric MOR. The main goal of this chapter is to introduce the concepts of "symmetry" and "structure" which play an important role in Lagrangian and Hamiltonian mechanics. Exploiting these symmetries and structures in MOR help to provide robustness and long-time stability of the reduced system.

2.1 Smooth Manifolds

Let $\mathcal{M}$ be a topological Hausdorff [44] set. Given a bijective map $x : U \to V \subset \mathbb{R}^m$, for some positive integer m and a neighborhood U of p, the pair (x, U) is called a *coordinate chart*. $\mathcal{M}$ is called a *smooth manifold* if there exists a set of coordinate charts $\{(x_\alpha, U_\alpha)\}_{\alpha \in I}$ such that $\{U_\alpha\}_{\alpha \in I}$ covers $\mathcal{M}$ and the mapping $x_{\alpha_2} \circ (x_{\alpha_1})^{-1} : \mathbb{R}^m \to \mathbb{R}^m$ is a C^∞ mapping, for any $\alpha_1, \alpha_2 \in I$. The integer m is the *dimension* of $\mathcal{M}$ if $x_\alpha(U_\alpha) \subset \mathbb{R}^m$ for all $\alpha \in I$. Throughout this thesis, we assume that there exists a global coordinate chart (x, U), i.e., U covers $\mathcal{M}$.

Vectors tangent to a manifold allow us to describe the local dynamics of an object moving on a smooth manifold. There are multiple ways to define such tangent vectors. The most intuitive way, however, uses curves defined on a smooth manifold.

A C^∞ mapping $\gamma : \mathbb{R} \to \mathcal{M}$ is called a curve on $\mathcal{M}$ passing through $p \in \mathcal{M}$ if $\gamma(t) = p$, for some $t \in \mathbb{R}$. Without loss of generality, we assume that γ passes through p at $t = 0$. We define an equivalence relation between smooth curves that pass through p as follows:

γ_1 and γ_2, two smooth curves on $\mathcal{M}$, are equivalent if

$$\frac{d}{dt}(x \circ \gamma_1)|_{t=0} = \frac{d}{dt}(x \circ \gamma_2)|_{t=0}, \tag{2.1}$$

for a coordinate chart x. It is easily checked that this definition is chart independent [1], i.e., the equivalent classes do not depend on the choice of the coordinate chart.

Definition 2.1. *A tangent vector v at a point $p \in \mathcal{M}$ is an equivalence class of curves on $\mathcal{M}$ that pass through p. The set of all tangent vectors at p is denoted by $T_p\mathcal{M}$ and is called the tangent space of $\mathcal{M}$ at point p.*

It is well known that the $T_p\mathcal{M}$ forms an m-dimensional linear vector space [1, 90]. It is shown in [1, 90] that the vector space $T_p\mathcal{M}$ is isomorphic to the vector space of directional derivatives, acting on the differentiable real-valued functions defined on an open neighborhood of p. Therefore, a basis for $T_p\mathcal{M}$ is often denoted as $\{\partial/\partial x_i\}_{i=1}^m$ and a vector $v \in T_p\mathcal{M}$ can be written as

$$\sum_{i=1}^{m} v_i \frac{\partial}{\partial x_i}, \tag{2.2}$$

where $v_i \in \mathbb{R}$ for $i = 1, \ldots, m$. Note that the symbol for the partial derivative is rather symbolic since the direction x_i is not defined explicitly on $T_p\mathcal{M}$. To compute the coefficients v_i, for $i = 1, \ldots, m$, let γ be a curve that is a representative of the equivalent class $v \in T_p\mathcal{M}$, the components of v in a given chart x are defined by the derivatives in Euclidean space of the curve $x \circ \gamma$, i.e.,

$$v_i = [\frac{d}{dt}(x \circ \gamma)]_i|_{t=0}.$$

To compute a tangent vector for a given chart x and a given smooth function $f : \mathcal{M} \to \mathbb{R}$ one considers

$$\sum_{i=1}^{m} v_i \frac{\partial}{\partial x_i}(f) := \sum_{i=1}^{m} v_i \frac{\partial}{\partial x_i}(f \circ x^{-1}), \tag{2.3}$$

where the partial derivatives that appear on the right hand side are the conventional partial derivative operators on a Euclidean space.

The dual space to $T_p\mathcal{M}$ is denoted by $T_p^*\mathcal{M}$ and is referred to as the *cotangent space*. The natural isomorphism between $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$ implies that the cotangent space is also an m-dimensional linear vector space. Given $\{\partial/\partial x_i\}$ as a basis for $T_p\mathcal{M}$, a dual basis for $T_p^*\mathcal{M}$ is a set of basis vectors $\{dx_i\}_{i=1}^m$ that satisfy

$$dx_i(\frac{\partial}{\partial x_j}) = \delta_{ij}, \quad i, j = 1, \ldots, m, \tag{2.4}$$

where δ_{ij} is the Kronecker's delta function.

To define a vector field on a manifold, we assign a tangent vector to every point on a manifold. Such an object belongs to a structure that, informally, is obtained by glueing the tangent space $T_p\mathcal{M}$ to every point $p \in \mathcal{M}$. This structure is referred to as the *tangent bundle*, denoted as $T\mathcal{M}$, and is defined as $T\mathcal{M} := \{(p, v) | p \in \mathcal{M},\ v \in T_p\mathcal{M}\}$.

Theorem 2.1. *[90] The tangent bundle $T\mathcal{M}$ is a smooth manifold.*

Proof. We define the projection operator $\pi : T\mathcal{M} \to \mathcal{M}$ as

$$\pi : (p, v) \to p. \tag{2.5}$$

It is easily checked that $\pi^{-1}(\{p\})$ is the m-dimensional linear vector space $T_p\mathcal{M}$. Now assume that (x, U) is a coordinate chart for $\mathcal{M}$, such that $p \in U$. We construct a coordinate chart $(\bar{x}, \pi^{-1}(U))$ for $T\mathcal{M}$ as follows

$$\begin{aligned} &\bar{x} : \pi^{-1}(U) \subset T\mathcal{M} \to \mathbb{R}^m \times \mathbb{R}^m, \\ &\bar{x} : (p, v) \to (x(p), v_1, \ldots, v_m), \end{aligned} \tag{2.6}$$

where $v_1, \ldots, v_m \in \mathbb{R}$ are components of v in $\mathbb{R}^m$. It can be seen that for two intersecting coordinate charts (x_α, U_α) and (x_β, U_β), the transition map $\bar{x}_\alpha \circ (\bar{x}_\beta)^{-1}$ is a C^∞ map. Thus, $T\mathcal{M}$ is a smooth manifold, and (2.6) suggests it is $2m$-dimensional. □

In a similar fashion, we may obtain a smooth manifold by gluing the cotangent space $T_p^*\mathcal{M}$ to the manifold $\mathcal{M}$ to obtain the *cotangent bundle*, denoted as $T^*\mathcal{M}$.

Definition 2.2. *Continuous injective mappings $X : \mathcal{M} \to T\mathcal{M}$ and $X^* : \mathcal{M} \to T^*\mathcal{M}$ are called a* vector field *and a* co-vector field *of $\mathcal{M}$, respectively.*

In other words, we assign a vector from $T_p\mathcal{M}$ to every point p on $\mathcal{M}$.

Definition 2.3. *Suppose that X is a vector field on a smooth manifold $\mathcal{M}$. A smooth curve $c : (a, b) \to \mathcal{M}$, $(a, b) \subset \mathbb{R}$, is called an* integral curve *of X if*

$$\frac{d}{dt}c(t) = X(c(t)), \quad \forall t \in (a, b). \tag{2.7}$$

Furthermore, we say that c passes through p if $c(t) = p$, for some $t \in (a, b)$.

Given a coordinate chart x, one can solve (2.7) for c. It is known from the theory of ordinary differential equations that (2.7) has a unique solution [102].

Definition 2.4. *The* flow *of X is a collection of maps $\varphi_t : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ such that the map $t \to \varphi_t(p)$ is an integral curve for some initial point $p \in \mathcal{M}$.*

Note that we have $\varphi_0 = id$, the identity map. Furthermore, the uniqueness of the integral curves implies the following important property of flows

$$\varphi_{t+s} = \varphi_t \circ \varphi_s. \tag{2.8}$$

In the study of MOR, transformations between vector spaces emerge naturally. In a more general setting, it is beneficial to restrict the study transformations between smooth manifolds. Later in this chapter, we discuss how some manifold structures can be preserved over such transformation, laying the foundation for geometric MOR.

Let $\mathcal{M}$ and $\mathcal{N}$ be an m-dimensional and an n-dimensional smooth manifolds, respectively. Furthermore, let $\phi : \mathcal{M} \to \mathcal{N}$ be a smooth mapping, i.e., if (x, U) is a coordinate chart for $\mathcal{M}$ and (y, V) is a coordinate chart for $\mathcal{N}$ such that $\phi(U) \cap V \neq \emptyset$, then the mapping $y \circ \phi \circ x^{-1}|_{U \cap \phi^{-1}(V)} : \mathbb{R}^m \to \mathbb{R}^n$ is C^∞. The *differential map* of ϕ at a point $p \in \mathcal{M}$, denoted by $T_p\phi$, is a map between the tangent spaces $T_p\mathcal{M}$ and $T_{\phi(p)}\mathcal{N}$ defined as

$$T_p\phi(v) = \frac{d}{dt}(\phi \circ \gamma(t))|_{t=0}, \tag{2.9}$$

for some tangent vector $v \in T_p\mathcal{M}$ and some curve γ in the equivalence class of v. It can be shown that $T_p\phi(v)$ only depends on v and not the choice of the curve γ [90]. The inverse function theorem [92] indicates that if $T_p\phi$ is a vector space isomorphism then there is a neighborhood U of p and a neighborhood V of $\phi(p)$, such that $\phi : U \to V$ is a diffeomorphism.

2.2 Tensors and Differential Forms

Quantities that appear in physics are often linearly dependent on the vectors and covectors that describe them. Examples of such a quantity are the measurement of the magnetic field with linearly independent directions of measurement [107], or the strength of resistance in dissipative fluid flows. Tensors and differential forms allow us to describe such objects, and provide the possibility to establish a relationship between vectors and vectors fields. In the following sections, we seek to align the flow of a system with respect to some vector field, in the context of symplectic geometry.

Definition 2.5. *A* differential k-form Ω *on a manifold* $\mathcal{M}$ *is a function*

$$\Omega_p : T_p\mathcal{M} \times \cdots \times T_p\mathcal{M} \ (k \textit{ times}) \to \mathbb{R}$$

that is multilinear

$$\Omega_p(v_1, \ldots, \alpha v_i + v_i', \ldots, v_k) = \alpha\Omega(v_1, \ldots, v_i, \ldots, v_k) + \Omega(v_1, \ldots, v_i', \ldots, v_k),$$

for $i = 1, \ldots, k$ *and some* $\alpha \in \mathbb{R}$*, and skew-symmetric*

$$\Omega_p(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -\Omega_p(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k),$$

for $i, j = 1, \ldots, k$ *and* $i \neq j$.

Definition 2.6. *A* (k, l)-tensor *on a manifold* $\mathcal{M}$ *is a multilinear (not necessarily skew-symmetric) function*

$$\Lambda_p : T_p^*\mathcal{M} \times \cdots \times T_p^*\mathcal{M}\ (k \text{ times}) \times T_p\mathcal{M} \times \cdots \times T_p\mathcal{M}\ (l \text{ times}) \to \mathbb{R}.$$

When the point the exact location of p is not important, we may drop the superscript p and write $\Omega = \Omega_p$ and $\Lambda = \Lambda_p$.

Given a coordinate chart x and a basis $\{e_1, \ldots, e_m\}$ for $T_p\mathcal{M}$, multilinearity of a k-form implies that

$$\Omega(v_1, \ldots, v_k) = \omega_{i_1,\ldots,i_k} v_1^{i_1} \ldots v_k^{i_k}, \quad \text{sum over } 1 \leq i_1, \ldots, i_k \leq m \tag{2.10}$$

where $\omega_{i_1,\ldots,i_k} = \Omega(e_{i_1}, \ldots, e_{i_k})$ and $v_l^{i_l}$ is the i_lth component of v_l with respect to the coordinate chart x. Therefore, any k-form is completely described by $\Omega(e_{i_1}, \ldots, e_{i_k})$, for $1 \leq i_1, \ldots, i_k \leq m$. A simple calculation confirms a similar results for (k, l)-tensors [107].

We now introduce some basic tensor operators, which allows the construction of higher order tensors and differential forms, from simpler building blocks.

Definition 2.7. *Let* Γ_1 *and* Γ_2 *be a* (k_1, l_l)*-tensor and a* (k_2, l_2)*-tensor, respectively. Their tensor product* $\Gamma_1 \otimes \Gamma_2$ *is a* $(k_1 + k_2, l_1 + l_2)$*-tensor defined as*

$$(\Gamma_1 \otimes \Gamma_2)(v_1^*, \ldots, v_{k_1+k_2}^*; w_1, \ldots w_{l_1+l_2}) = \\ \Gamma_1(v_1^*, \ldots, v_{k_1}^*; w_1, \ldots w_{l_1}) \cdot \Gamma_1(v_{k_1+1}^*, \ldots, v_{k_1+k_2}^*; w_{l_1+1}, \ldots w_{l_1+l_2}).$$

To be able to construct differential forms from $(0, k)$-tensors, we need an operator that skew-symmetrizes tensors. The *alternation operator*, is a tensor operator that achieves this and is defined as

$$\mathbf{A}(\Gamma)(v_1, \ldots, v_l) = \frac{1}{l!} \sum_{\pi \in S_l} \text{sgn}(\pi) \Gamma(v_{\pi(1)}, \ldots, v_{\pi(l)}). \tag{2.11}$$

Here Γ is a $(0, l)$-tensor, S_l is the permutation group of the set $\{1, \ldots, l\}$ and $\text{sgn}(\pi)$ returns 1 if π is an even permutation, and -1 if π is an odd permutation. It is easily checked that $\mathbf{A}(\Gamma)$ is skew-symmetric. Hence, $\mathbf{A}$ constructs a mapping from $\Lambda_{0,k}(\mathcal{M})$ to $\Lambda_k(\mathcal{M})$.

Note that a tensor product of two differential forms is not a differential form, since the skew-symmetry property will be lost. The *wedge product* allows us to construct higher order differential forms while preserving the skew-symmetry.

Definition 2.8. *Let Ω_1 and Ω_2 be a differential k_1-form and a k_2-form, respectively. The wedge product of Ω_1 and Ω_2 is a (k_1+k_2)-form defined as*

$$\Omega_1 \wedge \Omega_2 = \frac{k_1!+k_2!}{k_1!k_2!}\mathbf{A}(\Omega_1 \otimes \Omega_2). \tag{2.12}$$

It is well known that the wedge product is associative, bilinear and anti-commutative [69]. The following theorem states that any differential k-form can be written as a linear combination of wedge products of co-vectors. We refer the reader to [1] for the proof.

Theorem 2.2. *Any k-form Ω can be written locally as*

$$\Omega = \sum_{i_1<\cdots<i_k} \omega_{i_1,\ldots,i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \tag{2.13}$$

It is easily verified that $\Omega = 0$ implies $\omega_{i_1,\ldots,i_k} = 0$ for all $i_1 < \cdots < i_k$. Differential maps are often useful to transfer manifold structures from a known manifold to an unknown manifold. For example, for a mapping $\phi : \mathcal{M} \to \mathcal{N}$, $T_p\phi$ allows us to identify the tangent space of $\mathcal{N}$ at $\phi(p)$ using the tangent space of $\mathcal{M}$ at p. Similarly, we may use the differential maps to construct differential forms and tensors for unknown manifolds.

Definition 2.9. *Let $\phi : \mathcal{M} \to \mathcal{N}$ be a smooth manifold mapping and Ω be a differential k-form on $\mathcal{N}$. Then the* pull-back *of Ω with ϕ, is a k-form on $\mathcal{M}$, denoted by $\phi^*\Omega$, defined as*

$$(\phi^*\Omega)_p(v_1,\ldots,v_k) = \Omega_{\phi(p)}(T_p\phi(v_1),\ldots,T_p\phi(v_k)),$$

for any $p \in \mathcal{M}$ and $v_1 \ldots, v_k \in T_p\mathcal{M}$. In case ϕ is a diffeomorphism, the push-forward *operator, denoted as ϕ_*, is defined by $\phi_* = (\phi^{-1})^*$.*

Another basic operator on tensors and differential forms, is the *contraction* operator.

Definition 2.10. *Let Ω be a differential k-form and X be a smooth vector field on a smooth manifold $\mathcal{M}$. The contraction of Ω with respect to X is a $(k-1)$-form defined by*

$$(\mathbf{i}_X\Omega)_p(v_1,\ldots,v_{k-1}) = \Omega(X(p),v_1,\ldots,v_{k-1}).$$

The contraction operator is sometimes referred to as the *interior product*.

There are multiple ways to generalize the notion of differentiation to the manifold setting. The challenge in defining a unique derivative operator that is consistent with the conventional derivative operator in calculus, is that a general manifold setting does not provide an algebraic method to compare points on a manifold. To establish such a relation, the pull-back of a differential form can be used. The *Lie derivative* of points on a smooth manifold, employs this idea to generalize the notion of differentiation to the manifold setting.

Definition 2.11. *Let $\mathcal{M}$ be a smooth manifold and Ω be a differential k-form on $\mathcal{M}$. Given a vector field X on $\mathcal{M}$ with the flow map φ_t, the Lie derivative of Ω with respect to X is defined as*

$$\mathcal{L}_X\Omega = \lim_{t\to 0}\frac{1}{t}(\varphi_t^*\Omega - \Omega) = \frac{d}{dt}\varphi_t^*\Omega|_{t=0}. \tag{2.14}$$

Note that the flow map can be viewed as a mapping $\varphi_t : \mathcal{M} \to \mathcal{M}$, and thus, the differential map $T_p\varphi_t$ defines a mapping from $T_p\mathcal{M}$ to $T_{\varphi_t(p)}\mathcal{M}$. Therefore, the Lie derivative measures infinitesimal differences in Ω when evaluated at p and at $\varphi_t(p)$. The following theorem summarizes some basic properties of the Lie derivative.

Theorem 2.3. *Suppose that X is a smooth vector field defined on a smooth manifold $\mathcal{M}$ with φ_t being the flow of X. Furthermore, suppose that Ω is a differential k-form. The following statements hold:*

(a) for a smooth scalar function $f : \mathcal{M} \to \mathbb{R}$, $\mathcal{L}_X f = X \cdot f$, where $X \cdot f$ is the directional derivative of f along X.

(b) The Lie derivative formula

$$\frac{d}{dt}\varphi_t^*\Omega = \varphi_t^*\mathcal{L}_X\Omega.$$

(c) In case Ω is a time dependent differential form, i.e. $\Omega = \Omega_t$, we have

$$\frac{d}{dt}\varphi_t^*\Omega_t = \varphi_t^*\mathcal{L}_X\Omega + \varphi_t^*\frac{d}{dt}\Omega_t.$$

Proof. (a) follows from the definition of the Lie derivative. To show (b), we have

$$\frac{d}{dt}\varphi_t^*\Omega = \lim_{h\to 0}\frac{1}{h}(\varphi_{t+h}^*\Omega - \varphi_t^*\Omega) = \varphi_t^*(\lim_{h\to 0}\frac{1}{h}(\varphi_h^*\Omega - \Omega)) = \varphi_t^*\mathcal{L}_X\Omega, \tag{2.15}$$

where the second equality is due to the fact that $\varphi_{t+h} = \varphi_t \circ \varphi_h$. To show (c), assume that vectors $v_1 \ldots, v_k \in T_p\mathcal{M}$ at some point $p \in \mathcal{M}$ are provided . We define the function $\alpha(t,s) : \mathbb{R}^2 \to \mathbb{R}$ as $\alpha(t,s) = \varphi_t^*\Omega_s(v_1,\ldots,v_k)$. Therefore, $\frac{d}{dt}\varphi_t^*\Omega_t|_p$ is the directional

derivative of $\alpha(t,s)$ along the direction $(1,1)$. We have

$$\begin{aligned} D_{(1,1)}\alpha(t,s) = \partial_t\alpha(t,s) + \partial_s\alpha(t,s) &= \partial_t\varphi_t^*\Omega_s|_{t=s} + \varphi_t^*(\partial_s\Omega_s)|_{t=s} \\ &= \varphi_t^*\mathcal{L}_X\Omega_t + \varphi_t^*(\frac{d}{dt}\Omega_t). \end{aligned} \tag{2.16}$$

Here, $D_{(1,1)}$ is the conventional directional derivative and all terms are evaluated at the point p. In the second equality, we used statement (b) and the linearity of φ^*. Since the choice of the point p and vectors $v_1,\ldots,v_k$ is arbitrary, the formula holds independently. What is left, is to show that $\alpha(t,s)$ is differentiable. Given that Ω_t is continuously differentiable, we see from above that the partial derivatives of $\alpha(t,s)$ is continuous, thus, is differentiable. □

For a k-form Ω, the *exterior derivative* is a $(k+1)$-form $\mathbf{d}\Omega$, that captures the differential changes in Ω and for a given coordinate chart x is defined as

$$\mathbf{d}\Omega = \frac{\partial\omega_{i_1,\ldots,i_k}}{\partial x_j} dx_j \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}, \tag{2.17}$$

summing over j and $i_1 < \cdots < i_k$. Note that the above definition can be shown to be to be chart independent [93]. It can be checked checked that for $v_0, v_1 \ldots, v_k \in T_p\mathcal{M}$ and $p \in \mathcal{M}$ we have

$$\mathbf{d}\Omega_p(v_0,v_1,\ldots,v_k) = \sum_{i=1}^{k}(-1)^i \nabla\Omega \cdot v_i(v_0,\ldots,\mathring{v}_i,\ldots,v_k), \tag{2.18}$$

where $\nabla\Omega \cdot v_i = \sum_j \sum_{i_1<\cdots<i_k} v_i^j \partial\omega_{i_1,\ldots,i_k}/\partial x_j$. Here, $\mathring{v}_i$ implies that the vector v_i is dropped, and v_i^j is the jth component of v_i.

The following theorem expresses the Lie derivative of a differential form with respect to the exterior derivative and the contraction oprator. We state the theorem without a proof as it is legthly and out of context. We refer the reader to [1] for the full proof.

Theorem 2.4. *[1] (Cartan's Magic Formula) Let X be a smooth vector field and Ω be a differential k-form on a manifold $\mathcal{M}$. We have*

$$\mathcal{L}_X\Omega = \mathbf{d}\mathbf{i}_X\Omega + \mathbf{i}_X\mathbf{d}\Omega. \tag{2.19}$$

A differential k-form Ω is called *closed*, if $\mathbf{d}\Omega = 0$, and is called *exact* if there is a differential $(k-1)$-form Γ such that $\Omega = \mathbf{d}\Gamma$. The symmetry in the partial derivatives of smooth functions implies that the exterior derivative of any differential form is closed, i.e., $\mathbf{d}^2\Omega = 0$ for any differential k-form Ω. However, not all differential forms are exact. We close this section by showing the sufficient condition for a differential form to be exact.

Theorem 2.5. *(Poincaré Lemma) Let $\mathcal{M}$ be a smooth n-dimensional manifold with Ω a differential k-form, defined on $\mathcal{M}$. If Ω is closed then for any point $p \in \mathcal{M}$, there is a neighborhood U of p where $\Omega|_U$ is exact.*

Proof. Let (x, U) be a coordinate chart around $p \in \mathcal{M}$. Without loss of generality, we assume that $x(p) = 0$ and that the image of U under x contains an open ball around the origin. We show that the pull-back of Ω to any point on this ball is exact.

The condition $\mathbf{d}\Omega = 0$ implies that

$$\sum_l \sum_{j_1<\cdots<j_{k-1}} \frac{\partial \omega_{j_1,\ldots,j_k}}{\partial x_l} dx_l \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_k} = 0.$$

Note that the indices $l, j_1, \ldots, j_k$ are not in the proper order. It is easily checked that a reordering of the indices in the above expression yields

$$\sum_{n_1<\cdots<n_{k+1}} \lambda_{n_1,\ldots,n_{k+1}} dx_{n_1} \wedge \cdots \wedge dx_{n_{k+1}} = 0,$$

with

$$\lambda_{n_1,\ldots,n_{k+1}} = \sum_{m=1}^{k+1} (-1)^{m+1} \frac{\partial \omega_{n_1\ldots\mathring{n}_m\ldots n_{k+1}}}{\partial x_{n_m}},$$

which implies $\lambda_{n_1,\ldots,n_{k+1}} = 0$, for all $n_1 < \cdots < n_{k+1}$. Here $\mathring{n}_m$ implies that n_m is omitted. Now we construct a $(k-1)$-form Γ and claim that $\mathbf{d}\Gamma = \Omega$:

$$\Gamma_{(x_1,\ldots,x_n)} = \left(\int_0^1 t^{k-1} \omega_{j i_1\ldots,i_{k-1}}(tx_1, \ldots, tx_n) x_j \, dt \right) dx_{i_1} \wedge \cdots \wedge dx_{i_{k-1}}.$$

Taking the exterior derivative of Γ gives

$$\mathbf{d}\Gamma = \left(\int_0^1 t^{k-1} \frac{\partial}{\partial x_l} \left(\omega_{j i_1\ldots,i_{k-1}}(tx) x_j \right) dt \right) dx_l \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_{k-1}}.$$

Here tx denotes the point $(tx_1, \ldots, tx_n)$. Let $c_{l i_1\cdots_{k-1}}$ be the coefficients in this expression. Similar to the above, we can construct a reordering of the indices to obtain the coefficient of $dx_{j_1} \wedge \cdots \wedge dx_{j_k}$ for $j_1 < \cdots < j_k$ as

$$\sum_{m=1}^{k} (-1)^{m+1} c_{j_m j_1\ldots\mathring{j}_m\ldots j_k}.$$

It follows that

$$\begin{aligned}
&\sum_{m=1}^{k}(-1)^{m+1}c_{j_m j_1 \dots \mathring{j}_m \dots j_k} \\
&\quad = -\sum_{m=1}^{k}(-1)^m \int_0^1 t^{k-1}\frac{\partial}{\partial x_{j_m}}\left(\omega_{j j_1 \dots \mathring{j}_m \dots j_k}(tx)x_j\right) dt \\
&\quad = -\sum_{m=1}^{k}(-1)^m \int_0^1 t^{k-1}\left(tx_j\frac{\partial \omega_{j j_1 \dots \mathring{j}_m \dots j_k}}{\partial x_{j_m}}(tx) + \omega_{j_m j_1 \dots \mathring{j}_m \dots j_k}(tx)\right) dt \\
&\quad = \int_0^1 t^{k-1}\left(x_j t\sum_{m=1}^{k}(-1)^{m+1}\frac{\partial \omega_{j j_1 \dots \mathring{j}_m \dots j_k}}{\partial x_{j_m}}(tx) + k\omega_{j_1 \dots j_k}(tx)\right) dt \\
&\quad = \int_0^1 t^{k-1}\left(x_j kt\frac{\partial \omega_{j_1 \dots j_k}}{\partial x_j}(tx) + kt^{k-1}\omega_{j_1 \dots j_k}(tx)\right) dt \\
&\quad = \int_0^1 \frac{\partial}{\partial t}(t^k \omega_{j_1 \dots j_k}(tx))\, dt \\
&\quad = \omega_{j_1 \dots j_k}(x),
\end{aligned}$$

which implies $\mathrm{d}\Gamma = \Omega$. Here, we used the condition $\lambda_{n_1,\dots,n_{k+1}} = 0$ and the chain rule to obtain the result. □

2.3 Hamiltonian Systems on a Symplectic Manifold

It is often useful to describe small changes in a state of a system with respect to some potential or a vector field. Hamiltonian systems are systems where the changes in the state of the system are determined by a *Hamiltonian vector field*. Such systems appear in quantum physics, particle physics, celestial mechanics, cosmology, fluid mechanics, and classical mechanics. Conserved quantities, e.g. the system energy, are at the core of the dynamics of such systems. Consequently, the integral curve of these systems is aligned with the Hamiltonian vector field such that these quantities are conserved.

Differential forms are tools that allow us to align an integral curve with one or more vector fields. To study Hamiltonian systems, we therefore need to study basic features of differential 2-forms.

Definition 2.12. *Let $\mathcal{M}$ be a smooth manifold and $p \in \mathcal{M}$. The differential 2-form Ω_p is called* non-degenerate *if $\Omega_p(v_1, v_2) = 0$, for all $v_2 \in T_p\mathcal{M}$, implies that $v_1 = 0$. Ω is called non-degenerate, if Ω_p is non-degenerate for all $p \in \mathcal{M}$.*

If a non-zero vector $v \in T_p\mathcal{M}$ is given, then $\Omega_p^\flat(v) := \Omega_p(v, \cdot) : T_p\mathcal{M} \to \mathbb{R}$ can be viewed as a co-vector. Therefore, a non-degenerate Ω_p constructs an injective map $\Omega_p^\flat : T_p\mathcal{M} \to T_p^*\mathcal{M}$. When $\Omega_p^\flat$ is also surjective then it is said to be *strongly non-degenerate*.

Definition 2.13. *Let $\mathcal{P}$ be a smooth m-dimensional manifold and Ω be a closed, non-degenerate 2-form defined on $\mathcal{P}$. The pair $(\mathcal{P}, \Omega)$ is called a* symplectic manifold.

Note that the condition for Ω to be closed is indeed required to construct a well-defined symplectic manifold. In the following theorem, we see that closedness of the 2-form is required to locally approximate a symplectic manifold with a symplectic linear vector space. This is especially important in the context of symplectic MOR where we approximate a high dimensional symplectic manifold with a low dimensional linear vector space.

Theorem 2.6. *(Darboux' Theorem) Let $\mathcal{M}$ be a manifold and Ω_1 and Ω_2 be two strongly non-degenerate and closed 2-forms defined on $\mathcal{M}$, such that $\Omega_1 = \Omega_2$, at some $p \in \mathcal{M}$. Then there are neighborhoods U and V of p such that the mapping $\varphi : U \to V$ is a diffeomorphism with $\varphi^*\Omega_2 = \Omega_1$*

Proof. The idea is to construct a family of continuously varying 2-forms:

$$\Omega_t = (1-t)\Omega_0 + t\Omega_1 = \Omega_0 + t\Omega,$$

where $\Omega = \Omega_1 - \Omega_0$. Now we would like to find a smooth vector field X_t with the flow φ_t such that $\frac{d}{dt}\varphi_t^*\Omega_t = 0$. We construct U so small that Ω_t is strongly non-degenerate. This can be done since $\Omega_0 = \Omega_1$ and is constant at p, so the compactness of $[0,1]$ implies that there is an open ball around p such that Ω_t is strongly non-degenerate for all $t \in [0,1]$. We have

$$\begin{aligned}\frac{d}{dt}\varphi_t^*\Omega_t &= \varphi_t^*\mathcal{L}_{X_t}\Omega_t + \varphi_t^*\frac{d}{dt}\Omega_t \\ &= \varphi_t^*(\mathbf{d}i_{X_t}\Omega_t + i_{X_t}\mathbf{d}\Omega_t) + \varphi_t^*\Omega = \varphi_t^*(\mathbf{d}i_{X_t}\Omega_t + \Omega).\end{aligned}$$

Here we used the Lie derivative formula for time dependent differential forms, Cartan's magic formula, closedness of Ω_t, and the linearity of pull-back operator. Since Ω is a closed form and strongly non-degenerate, we can apply the Poincaré lemma 2.5 to obtain $\Omega = \mathbf{d}\Gamma$ for some 1-form Γ. The above expression becomes

$$\frac{d}{dt}\varphi_t^*\Omega_t = \varphi_t^*(\mathbf{d}(i_{X_t}\Omega_t + \Gamma)).$$

Therefore, it is sufficient to define X_t to be the vector field associate to the relation $i_{X_t}\Omega_t = -\Gamma$. Note that the non-degeneracy of Ω_t in U guarantees the uniqueness of X_t. Thus, $\varphi_1^*\Omega_1 = \varphi_0^*\Omega_0 = \Omega_0$ □

Corollary 2.7. *Let $(\mathcal{P}, \Omega)$ be a symplectic manifold. There is a local coordinate chart (x, U) around each point $z \in \mathcal{P}$ for which Ω is constant.*

Proof. Fix a point $z_0 \in \mathcal{P}$. Take $\Omega_0 = \Omega$ and $\Omega_1 = \Omega_{z_0}$, i.e., Ω_1 is a constant differential

form $\Omega|_{z=z0}$. The rest follows from the proof of Darboux' theorem. The constructed flow φ_t provides a coordinate chart that transforms Ω into the constant form Ω_1. □

Corollary 2.8. *Let $(\mathcal{P}, \Omega)$ be a finite-dimensional symplectic manifold. Then $\mathcal{P}$ is even dimensional and we can find a local coordinate chard (x, U) with $x = (e_1, \ldots, e_n, f_1, \ldots, f_n)$ around each point $z \in \mathcal{P}$ such that*

$$\Omega = \sum_{i=1}^{n} de_i \wedge df_i. \tag{2.20}$$

This local coordinate chart is referred to as the canonical basis.

Proof. (The symplectic Gram-Schmidt) Suppose that (x, U) is a local coordinate chart around z provided by the Darboux' theorem such that Ω_z is constant, i.e., $\Omega_z = \Omega$. Let $e_1 \in T_z\mathcal{P}$ be a nonzero tangent vector. Non-degeneracy of Ω implies that there is a vector $f_1 \in T_z\mathcal{P}$ such that $\Omega(e_1, f_1) = c_1 \neq 0$. We can swap and scale e_1 and f_1 to guarantee that $c_1 = 1$. Let $E_1 = \text{span}\{e_1, f_1\}$ and $E_2 = \{v \in T_z\mathcal{P} | \Omega(v, e) = 0, \forall e \in E_1\}$. It is easily verified that $E_1 \cap E_2 = \emptyset$. Furthermore, for any $v \in T_z\mathcal{P}$, $v - \bar{v} \in E_2$ where $\bar{v} = -\Omega(v, f_1)e_1 + \Omega(v, e_1)f_1 \in E_1$, therefore $T_z\mathcal{P} = E_1 \oplus E_2$. Finally, since Ω is non-degenerate, it is also non-degenerate on E_2 as a subspace of $T_z\mathcal{P}$. Thus, we can continue inductively. Since $\mathcal{P}$ is finite dimensional this process ends, e.g., after n steps. Furthermore, the sequence of basis vectors $A = \{e_1, \ldots, e_n, f_1, \ldots, f_n\}$ forms a basis for $T_z\mathcal{P}$. In this basis, Ω takes the *canonical* form

$$\Omega(u, v) = \xi^T \mathbb{J}_{2n} \eta, \quad \mathbb{J}_{2n} = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix}. \tag{2.21}$$

Here, $\xi, \eta \in \mathbb{R}^{2n}$ with $\xi_i = dx_i(u)$ and $\eta_i = dx_i(v)$, $i = 1, \ldots, 2n$ for any $u, v \in T_z\mathcal{P}$. This is the matrix notation of the form in (2.20). □

Due to non-degeneracy of Ω, vectors $f_1, \ldots, f_n$ can be interpreted as co-vectors such that $\Omega^\flat(e_i) = f_i$, for $i = 1, \ldots, n$. By abusing the notation, the co-vector property of these vectors is denoted by $f_1^*, \ldots, f_n^*$.

Corollary 2.9. *Let $(\mathcal{P}, \Omega)$ be a symplectic manifold. There is a neighborhood U around each point $z \in \mathcal{P}$ where U is diffeomophic to a linear vector space V where Ω is constant on V.*

Proof. The proof is a direct consequence of the Darboux' theorem and Corollary 2.8. □

Definition 2.14. *Let $(\mathcal{P}_1, \Omega_1)$ and $(\mathcal{P}_2, \Omega_2)$ be two symplectic manifolds. The transformation $\varphi : \mathcal{P}_1 \to \mathcal{P}_2$ is called a* symplectic transformation *if*

$$\varphi^* \Omega_2 = \Omega_1. \tag{2.22}$$

Now we are ready to define Hamiltonian systems on a symplectic manifold.

Definition 2.15. *Let $(\mathcal{P}, \Omega)$ be a symplectic manifold. We refer to a vector field X_H as a Hamiltonian vector field, if we can find a real function $H : \mathcal{P} \to \mathbb{R}$ such that*

$$i_{X_H} \Omega = \mathbf{d}H. \tag{2.23}$$

We call H the Hamiltonian function. *In this case, the equations of evolution is given by*

$$\dot{z} = X_H(z), \tag{2.24}$$

and is referred to as Hamilton's equation of evolution.

When $\mathcal{P}$ is $2n$-dimensional and a canonical basis is provided, Hamilton's equation take the form

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad i = 1, \ldots, n. \tag{2.25}$$

The following proposition shows that the flow of a Hamiltonian vector field is a symplectic map.

Proposition 2.10. *Let φ_t be the flow of a Hamiltonian vector field X_H. Then $\varphi_t : \mathcal{P} \to \mathcal{P}$ is a symplectic transformation.*

Proof. We have

$$\frac{d}{dt} \varphi_t^* \Omega = \varphi_t^* \mathcal{L}_{X_H} \Omega = \varphi_t^* (i_{X_H} \mathbf{d}\Omega + \mathbf{d} i_{X_H} \Omega) = \varphi_t^* (i_{X_H} \mathbf{d}\Omega + \mathbf{d}^2 H).$$

However, closedness of Ω implies $\mathbf{d}\Omega = 0$. Furthermore, since any exact differential form is closed, then $\mathbf{d}^2 \Omega = 0$. Thus, $\frac{d}{dt} \varphi_t^* \Omega = 0$ and $\varphi_t^* \Omega = \varphi_0^* \Omega = \Omega$. □

Corollary 2.11. *The Hamiltonian is conserved along the Hamiltonian flow.*

Proof. It follows that

$$\begin{aligned} \frac{d}{dt} H(\varphi_t(z)) = \mathbf{d}H \left(\frac{d}{dt} \varphi_t(z) \right) = \mathbf{d}H \left(X_H(z) \right) &= i_{X_H(z)} \Omega(X_H(z)) \\ &= \Omega(X_H(z), X_H(z)) = 0. \end{aligned}$$

Therefore, $H \circ \varphi_t$ is constant in time. □

2.4 Hamiltonian Systems on a Symplectic Linear Vector Space

In the symplectic MOR, the flow of a Hamiltonian system is projected on a low dimensional symplectic linear vector space. Therefore, it is beneficial to investigate symplectic linear vector spaces. In this section we assume that a symplectic manifold $(\mathcal{P}, \Omega)$ is also a linear vector space, i.e., we identify both $\mathcal{P}$ and $T_z\mathcal{P}$ with a linear vector space $\mathcal{Z}$.

Definition 2.16. *Let $\mathcal{Z}$ be a finite dimensional linear vector space with Ω a constant 2-form defined on $\mathcal{Z}$. The pair $(\mathcal{Z}, \Omega)$ is called a symplectic linear vector space if Ω is non-degenerate.*

Note that Corollary 2.8 indicates that $\mathcal{Z}$ is even-dimensional. Furthermore, the symplectic Gram-Schmidt process can construct a canonical basis, in which $\Omega(u, v) = \xi^T \mathbb{J}_{2n} \eta$, where $u, v \in \mathcal{Z}$, $\xi, \eta \in \mathbb{R}^{2n}$ are the expansion coefficients of u and v and $\mathbb{J}_{2n}$ is defined in (2.21). However, in a non-canonical basis, the symplectic form Ω takes the form $\Omega(u, v) = \xi^T J_{2n} \eta$, where J_{2n} is a full rank and skew-symmetric matrix. Therefore, when the matrix form of Ω is discussed, we may refer to Ω as $\mathbb{J}_{2n}$ or J_{2n}, depending on the coordinate chart.

A symplectic linear vector space can be equipped with an inner product to form a inner product space. However, a symplectic basis is not necessarily well-conditioned (or orthonormal) with respect to a general inner-product. The following proposition provides a natural inner product defined on a symplectic linear vector space that carries the conditioning of a canonical basis.

Proposition 2.12. *Let $(\mathcal{Z}, \Omega)$ be a $2n$-dimensional symplectic linear vector space. Furthermore, let $A = \{e_1, \ldots, e_n, f_1, \ldots, f_n\}$ be a canonical basis for $\mathcal{Z}$ with respect to Ω. Then there exists an inner product operator $<,>: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$, such that A is orthonormal.*

Proof. Consider the linear mapping $\Omega^\flat : \mathcal{Z} \to \mathcal{Z}^*$ and let e_i^* and f_i^* be the image of e_i and f_i under $\Omega^\flat$, respectively for $i = 1, \ldots, n$. Non-degeneracy of $\Omega^\flat$ indicates that e_i^* and f_i^* are unique and non-zero, and that $A^* = \{e_1^*, \ldots, e_n^*, f_1^*, \ldots, f_n^*\}$ forms a basis for $\mathcal{Z}^*$. We construct a linear map $T : \mathcal{Z}^* \to \mathcal{Z}$ by prescribing its action on elements of A^* as

$$T(e_i^*) = f_i, \quad T(f_i^*) = -e_i, \quad i = 1, \ldots, n.$$

Note that T is bijective. Now define the operator $<,>: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ as

$$< u, v >= T_u^{-1}(v),$$

where $T_u^{-1} = T^{-1}(u) \in \mathcal{Z}^*$. It is straight forward to check that $<,>$ is symmetric, bilinear and positive-definite. Therefore, it defines an inner-product on $\mathcal{Z}$. Furthermore, we have

$$< e_i, e_j >= \delta_{ij}, \quad < f_i, f_j >= \delta_{ij}, \quad < e_i, f_j >= 0, \quad i, j = 1, \ldots, n.$$

□

Corollary 2.13. *In a canonical $A = \{e_i, f_i\}_{i=1}^n$, the inner product constructed in Proposition 2.12 is the Euclidean inner product.*

Proof. Let u, v be a two vectors in $\mathcal{Z}$, and v^* be the image of v under $\Omega^\flat$. Let $\xi, \eta \in \mathbb{R}^{2n}$, be the expansion coefficients vector of u and v in the basis of A, respectively. Recall that $\Omega(u, v) = \xi^T \mathbb{J}_{2n} \eta$ in a canonical coordinate system.

We have

$$\begin{aligned} T^{-1}(u) = T^{-1}\left(\sum_i (\alpha_i e_i + \alpha_i' e_i)\right) &= \sum_i \left(\alpha_i T^{-1}(e_i) + \alpha_i' T^{-1}(f_i)\right) \\ &= \sum_i \left(-\alpha_i f_i^* + \alpha_i' e_i^*\right), \end{aligned} \tag{2.26}$$

where α_i and α_i' are the ith and the $(n+i)$th element of ξ, respectively. It follows

$$\begin{aligned} < u, v >= T_u^{-1}(v) &= \left(\sum_i \left(-\alpha_i f_i^* + \alpha_i' e_i^*\right)\right)\left(\sum_j \left(\beta_i e_i + \beta_i' f_i\right)\right) \\ &= \sum_i \alpha_i \beta_i + \alpha_i' \beta_i' = \xi^T \eta. \end{aligned} \tag{2.27}$$

Here, β_i and β_i' are the ith and the $(n+i)$th element of η, respectively. Furthermore, in the last step, we used the fact that $e_i^* f_j = \Omega(e_i, f_j) = \delta_{ij}$, for $i, j = 1, \ldots, n$, and the skew-symmetry of Ω. Furthermore, If we represent a co-vector as a transpose of a vector, (2.26) implies that then then expansion coefficients of u^* is given by $\mathbb{J}_{2n}\xi$. □

An inner product defined in Proposition 2.12 is especially useful since one can switch between computing the 2-form Ω and the inner production $<,>$, as needed. However, when dealing with a subspace of $\mathcal{Z}$, further structures are required to guarantee that such inner-product exists.

Definition 2.17. *[1] Let $(\mathcal{Z},\Omega)$ be a symplectic linear vector space and $\mathcal{E}\subset\mathcal{Z}$ be a subspace. The* symplectic complement *of $\mathcal{E}$, referred to as $\mathcal{E}^\perp$, is a linear subspace defined as*

$$\mathcal{E}^\perp=\{a\in\mathcal{Z}|\Omega(a,b)=0,\ \textit{for all}\ b\in\mathcal{E}\}.$$

Furthermore, we call $\mathcal{E}$ an isotropic subspace *if $\mathcal{E}\subset\mathcal{E}^\perp$, a* symplectic subspace *if $\Omega|_{\mathcal{E}\times\mathcal{E}}$ is non-degenerate, and a* Lagrangian subspace *if both $\mathcal{E}$ and $\mathcal{E}^C$ are isotropic, where $\mathcal{E}^C$ is vector space complement of $\mathcal{E}$*[1].

Proposition 2.14. *[1] Let $(\mathcal{Z},\Omega)$ be a symplectic linear vector space and $\mathcal{E}\subset\mathcal{F}$ is a linear subspace. Then the following are equivalent:*

(a) $\mathcal{E}$ is a Lagrangian subspace.

(b) $\mathcal{E}=\mathcal{E}^\perp$.

(c) $\mathcal{E}$ is isotropic and $dim(\mathcal{E})=\frac{1}{2}dim(\mathcal{Z})$.

Proof. We first prove that (a) implies (b). Since a Lagrangian subspace is isotropic then $\mathcal{E}\subset\mathcal{E}^\perp$. Let $F=E^C$ and take Take $v\in\mathcal{E}^\perp$. It is easily verified that $\mathcal{Z}=\mathcal{E}\oplus\mathcal{F}$. Therefore, $v=e+f$, where $e\in\mathcal{E}$ and $f\in\mathcal{F}$. Since $\mathcal{E}$ is isotropic then $\Omega(e,u)=0$ for all $u\in\mathcal{E}$. As $v\in\mathcal{E}^\perp$, it follows that for any $u\in E$

$$0=\Omega(v,u)=\Omega(e,u)+\Omega(f,u)=\Omega(f,u).$$

But since $\mathcal{F}$ is isotropic then $\Omega(f,u)=0$ for all $u\in\mathcal{F}$. This implies that $\Omega(f,u)=0$ for all $u\in\mathcal{E}\cup\mathcal{F}=\mathcal{Z}$. Non-degeneracy of Ω implies that $f=0$ and $v=e\in\mathcal{E}$. Therefore $\mathcal{E}^\perp\subset\mathcal{E}$.

Now we show that (b) implies (c). Consider the mapping $\Omega^\flat:\mathcal{Z}\to\mathcal{Z}^*$. Note that $\mathcal{E}^\perp$ is in the kernel of $\Omega^\flat|_\mathcal{E}:\mathcal{E}\to(\mathcal{Z}\backslash\mathcal{E}^\perp)^*$. Since $\Omega^\flat$ is injective, $\dim(\mathcal{E})\le\dim(\mathcal{Z})-\dim(\mathcal{E}^\perp)$. Furthermore, $\Omega^\flat|_\mathcal{E}$ can be viewed as a mapping from $\mathcal{Z}$ to $\mathcal{E}^*$ where the kernel is exactly $\mathcal{E}^\perp$. Therefore, $\dim(\mathcal{E})\ge\dim(\text{range}(\Omega^\flat|_\mathcal{E}))=\dim(\mathcal{Z})-\dim(\mathcal{E}^\perp)$. The two inequalities imply $\dim(\mathcal{Z})=\dim(\mathcal{E})+\dim(\mathcal{E}^\perp)$ which, together with (b), provide the result.

Finally, we show that (c) implies (a). Note that from the above results we conclude that $\dim(\mathcal{E})=\dim(\mathcal{E}^\perp)$. But since $\mathcal{E}\subset\mathcal{E}^\perp$, then $\mathcal{E}=\mathcal{E}^\perp$. Now we show that $\mathcal{F}=\mathcal{E}^C$ is isotropic. Take $f_1\notin\mathcal{E}$, and define $\mathcal{F}_1=\text{span}\{f_1\}$. Since $\mathcal{E}\cap\mathcal{F}_1=\emptyset$, it follows[2] that $\mathcal{E}+\mathcal{F}_1^\perp=\mathcal{Z}$. Therefore, we can pick $f_2\in\mathcal{F}_1^\perp$ such that $f_2\notin\mathcal{E}$, and define $\mathcal{F}_2=\text{span}\{f_2\}+\mathcal{F}_1$. We continue this process inductively to construct $\mathcal{F}_n$ such that

[1] E^C is a vector space complement of E if $\mathcal{Z}=\mathcal{E}\oplus\mathcal{E}^C$.

[2] For subsets E and F we have $(E\cap F)^\perp=(E^{\perp\perp}\cap F^{\perp\perp})^\perp=(F^\perp+E^\perp)^{\perp\perp}=F^\perp+E^\perp$ [1].

$\mathcal{Z} = \mathcal{E} \oplus \mathcal{F}_n$. It follows that

$$\begin{aligned}\mathcal{F}_n = \text{span}\{f_1, \ldots, f_n\} &\subset \text{span}(f_1)^\perp \cap \cdots \cap \text{span}(f_n)^\perp \\ &= (\text{span}(f_1) + \cdots + \text{span}(f_n))^\perp = \mathcal{F}_n^\perp.\end{aligned}$$

The second step uses the fact that each f_i is taken from $\mathcal{F}_{i-1}^\perp$. Therefore $\mathcal{F} = \mathcal{F}_n$ is isotropic. This completes the proof. □

Proposition 2.14 suggests that a Lagrangian subspace is a maximal isotropic subspace. We are now ready to state the following theorem.

Theorem 2.15. *[1] Let $(\mathcal{Z}, \Omega)$ be a $2n$-dimensional symplectic linear vector space. Furthermore, suppose that $\mathcal{X} \subset \mathcal{Z}$ is a $2k$-dimensional linear subspace with $k < n$. If the pair $(\mathcal{X}, \Omega|_\mathcal{X})$ contains a Lagrangian subspace, then $(\mathcal{X}, \Omega|_\mathcal{X})$ is also a symplectic subspace.*

Proof. We show that $\Omega|_\mathcal{X}$ is non-degenerate on $\mathcal{X}$. Let $\mathcal{E}$ be a Lagrangian subspace of $\mathcal{X}$. Since $\mathcal{F} = \mathcal{E}^C$ and $\dim(F) = \dim(E) = \dim(\mathcal{X})/2$ then, by Proposition 2.14 (c), $\mathcal{F}$ is also a Lagrangian subspace. Furthermore, Proposition 2.14 implies that $\mathcal{E} = \mathcal{E}^\perp$ and $\mathcal{F} = \mathcal{F}^\perp$. Now assume that there is $u \in \mathcal{X}$ such that $\Omega|_\mathcal{X}(u, v) = 0$ for all $v \in \mathcal{X}$. Then $v \in \mathcal{E}^\perp = \mathcal{E}$ and $v \in \mathcal{F}^\perp = \mathcal{F}$. Therefore $v \in \mathcal{E} \cap \mathcal{F} = \{0\}$. □

Once a linear Lagrangian subspace of a symplectic linear vector space is identified, one can induce a reduced symplectic form $\Omega|_\mathcal{X}$. This is particularly important in the context of MOR where a low dimensional symplectic reduced vector space is constructed. The induced symplectic form $\Omega|_\mathcal{X}$ can then be used to construct an ortho-symplectic basis on the low dimensional symplectic subspace.

2.5 Symplectic Integration of Hamiltonian Systems

In section 2.3 we saw two intrinsic feature of Hamiltonian systems: conservation of the Hamiltonian, expressed in Corollary 2.11, and the symplecticity of the Hamiltonian flow, discussed in Proposition 2.10. Since the exact flow of a Hamiltonian system is not generally available, the approximation of this flow has been a main topic of study, offen referred to as geometric numerical integration [51, 18]. However, there are no known methods that can preserve both the conservation of the Hamiltonian and the symplecticity of the flow for a general Hamiltonian system. *Symplectic numerical integration*, is a class of methods that approximate the flow of a Hamiltonian systems while preserving the symplecticity of the flow. Although these methods tend to violated the conservation of the Hamiltonian, robustness over long-time integration makes them the method of choice in wide range of application, e.g. molecular dynamics

simulation [35] and celestial mechanics [96]. In this section we investigate the theory behind symplectic integration and introduce some common symplectic numerical integrators. We assume, in this section, that $(\mathcal{Z}, \Omega)$ is a symplectic linear vector space (often $\mathcal{Z} = \mathbb{R}^{2n}$) with a constant Ω. Furthermore, we assume a canonical coordinate system, in which Ω takes the matrix form $\mathbb{J}_{2n}$. We emphasize this by writing $\Omega = \Omega_{\mathbb{J}_{2n}}$.

Proposition 2.16. *Let $(\mathbb{R}^{2n}, \Omega_{\mathbb{J}_{2n}})$ be a $2n$-dimensional symplectic linear vector space and let $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be a smooth symplectic diffeomorphism, then*

$$(\psi')^T \mathbb{J}_{2n} \psi' = \mathbb{J}_{2n}. \tag{2.28}$$

Here, ψ' is the Jacobian matrix of ψ.

Proof. Let $v_1, v_2 \in \mathbb{R}^{2n}$. Symplecticity of ψ implies that

$$\Omega_{\mathbb{J}_{2n}}(v_1, v_2) = \psi^* \Omega_{\mathbb{J}_{2n}}(v_1, v_2) = \Omega_{\mathbb{J}_{2n}}(T_z\psi(v_1), T_z\psi(v_2)) = \Omega_{\mathbb{J}_{2n}}(\psi' \cdot v_1, \psi' \cdot v_2),$$

Therefore,

$$v_1^T \mathbb{J}_{2n} v_2 = v_1^T (\psi')^T \mathbb{J}_{2n} \psi' v_2.$$

Since this is true for any $v_1, v_2 \in \mathbb{R}^{2n}$, (2.28) must hold. □

Definition 2.18. *When ψ is a linear transformation, i.e., $\psi(z) = Az$ with $A \in \mathbb{R}^{2n \times 2n}$,* (2.28) *takes the form*

$$A^T \mathbb{J}_{2n} A = \mathbb{J}_{2n}. \tag{2.29}$$

In this case the matrix A is referred to as a symplectic matrix.

To explain the intuition behind a symplectic transformation, we use the geometric interpretation of a symplectic transformation form [51]. Suppose that $f : U \to \mathbb{R}^{2n}$, with U a compact subset of $\mathbb{R}^2$, a 2-dimensional smooth surface in $\mathbb{R}^{2n}$. The manifold $S = f(U)$ can be viewed as the union of infinitesimal parallelograms spanned by vectors $\partial f/\partial x$ and $\partial f/\partial y$. We now define an area operator that recovers the area of the surface by summing over the area of all parallelograms:

$$\text{Area}(M) := \int_U \Omega_{\mathbb{J}_{2n}}(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})\, dxdy. \tag{2.30}$$

Given a symplectic diffeomorphism $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, we have

$$\text{Area}(\psi(M)) = \int_U \Omega_{\mathbb{J}_{2n}}(\frac{\partial(\psi \circ f)}{\partial x}, \frac{\partial(\psi \circ f)}{\partial y})\, dxdy = \text{Area}(M).$$

Therefore, the area, defined in (2.30), is preserved under a symplectic transformation. We extend this idea to the volume of a compact subset of a hypersurface by introducing the volume form

$$V = \frac{-1^{n(n-1)/2}}{n!} \underbrace{\Omega_{\mathbb{J}_{2n}} \wedge \cdots \wedge \Omega_{\mathbb{J}_{2n}}}_{n \text{ times}}.$$

This differential $2n$-form is referred to as the *Liouville volume* [69] form and the coefficient $-1^{n(n-1)/2}/n!$ is chosen such that V takes the form

$$V = de_1 \wedge \cdots \wedge de_n \wedge df_1 \wedge \cdots \wedge df_n. \tag{2.31}$$

in a canonical coordinate. Given vectors $v_1 \ldots, v_{2n} \in \mathcal{Z}$, the Liouville volume form computes the volume of a hyper dimensional parallelepiped spanned by $v_1 \ldots, v_{2n}$. The symplecticity of $\Omega_{\mathbb{J}_{2n}}$, implies that for a symplectic diffeomorphism $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $\psi^* V = V$ and, thus, a symplectic transformation preserves the volume with respect to the Liouville volume form.

Definition 2.19. *Let φ_t be the flow of a smooth vector field X defined on $\mathbb{R}^{2n}$. The mapping $\Phi : \mathbb{R} \times \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is a one step method if $\Phi_h(z_k) = z_{k+1} \approx \varphi_h(z_k)$. Furthermore, Φ is called a* symplectic one step method *if φ_t is a Hamiltonian flow and Φ is a symplectic transformation.*

Definition 2.20. *A one step method has order p if for all smooth φ_t we have*

$$\Phi_h(z_0) - \varphi_h(z_0) = \mathcal{O}(h^{p+1}), \quad z_0 \in \mathbb{R}^{2n}. \tag{2.32}$$

Definition 2.21. *Φ_h^* is called the* adjoint *of a one step method Φ_h, when it is the inverse map of Φ_h with the reversed time step $-h$, i.e.,*

$$\Phi_h^* = \Phi_{-h}^{-1}. \tag{2.33}$$

Since the exact flow of a Hamiltonian system is generally unknown, we approximate it with a stepping method, e.g. a one step method. The symplecticity of the Hamiltonian flow contributes to many geometrical symmetries, e.g. preservation of volume as discussed above. Thus, it is natural to expect a one step method to share this property in a numerical evaluation of Hamiltonian systems. Note that a general stepping schemes is not volume preserving and can result in a qualitatively wrong approximation of a Hamiltonian flow [51]. On the other hand, area preservation of symplectic time stepping schemes ensures excellent long-time behaviour [51, 18].

Definition 2.22. *Let $H(q,p)$ be a smooth Hamiltonian where $q, p \in \mathbb{R}^n$ are vectors containing the coefficients of a vector in a canonical basis $A = \{e_i, f_i\}_{i=1}^n$. The one step*

method

$$q_{k+1} = q_k + h \cdot \nabla_p H(q_{k+1}, p_k), \quad p_{k+1} = p_k - h \cdot \nabla_q H(q_{k+1}, p_k) \tag{2.34}$$

is called the symplectic Euler *scheme and approximates the flow of the Hamiltonian system* $\dot{z} = \mathbb{J}\nabla_z H$*, with* $z = (q,p)^T$*. Here,* ∇_x *indicates the gradient operator with respect to the vector* x*.*

Theorem 2.17. *[51] The symplectic Euler methods is a symplectic one step method of order 1.*

Proof. Let $\Phi_h(q_k, p_k) = (q_{k+1}, p_{k+1})$ be the one step method representing the symplectic Euler method. It follows

$$\begin{pmatrix} I_n + hH_{pq}^T & 0 \\ -hH_{qq} & I \end{pmatrix} \Phi_h' = \begin{pmatrix} I & -hH_{pp} \\ 0 & I + hH_{pq} \end{pmatrix},$$

where H_{yx} is an $n \times n$ matrix containing $\partial^2 H/(\partial y_j \partial x_i)$ in its i-th row and j-th column, evaluated at q_{k+1}, p_k. It is straightforward to check that $(\Phi')^T \mathbb{J}_{2n} \Phi' = \mathbb{J}_{2n}$. Furthermore, a Taylor expansion verifies that this method is of order 1 [51]. □

The adjoint of the symplectic Euler is also a symplectic one step method of order 1 [51], and is given by

$$q_{k+1} = q_k + h \cdot \nabla_p H(q_k, p_{k+1}), \quad p_{k+1} = p_k - h \cdot \nabla_q H(q_k, p_{k+1}). \tag{2.35}$$

It is straightforward to show that the composition of symplectic time stepping methods is again a symplectic method. Thus, one way to construct higher order methods is to compose one step methods.

Theorem 2.18. *The Störmer-Verlet time stepping scheme, given by*

$$\begin{aligned} q_{k+1/2} &= q_n + \frac{h}{2} \cdot \nabla_p H(q_{k+1/2}, p_k), \\ p_{k+1} &= p_k - \frac{h}{2} \cdot (\nabla_q H(q_{k+1/2}, p_k) + \nabla_q H(q_{k+1/2}, p_{k+1/2})), \\ q_{k+1} &= q_{k+1/2} + \frac{h}{2} \cdot \nabla_p H(q_{k+1/2}, p_{k+1/2}), \end{aligned} \tag{2.36}$$

or

$$\begin{aligned} p_{k+1/2} &= p_n - \frac{h}{2} \cdot \nabla_p H(q_k, p_{k+1/2}), \\ q_{k+1} &= q_k + \frac{h}{2} \cdot (\nabla_p H(q_k, p_{k+1/2}) + \nabla_p H(q_{k+1/2}, p_{k+1/2})), \\ p_{k+1} &= p_{k+1/2} - \frac{h}{2} \cdot \nabla_p H(q_{k+1/2}, p_{k+1/2}), \end{aligned} \tag{2.37}$$

is a symplectic time stepping method of order 2.

Proof. One checks that (2.36) and (2.37) are obtained by composing two symplectic Euler methods with step $h/2$ as $\Phi_{h/2} \circ \Phi^*_{h/2}$ and $\Phi^*_{h/2} \circ \Phi_{h/2}$, respectively. This implies that the Störmer-verlet scheme is of order 2. □

Theorem 2.19. *The implicit midpoint scheme for a Hamiltonian system*

$$z_{k+1} = z_k + h \cdot J_{2n} \nabla_z H(\frac{1}{2}(z_{k+1} + z_k)), \tag{2.38}$$

is a symplectic scheme of order 2.

Proof. Let $\Phi_h(z_k) = z_{k+1}$ be the one step method corresponding to the implicit midpoint rule. It follows

$$(I - \frac{h}{2} J_{2n} \nabla_z^2 H)\Phi'_h = (I + \frac{h}{2} J_{2n} \nabla_z^2 H),$$

where ∇^2 indicates the Hessian operator. It is straight forward to show that $(\Phi'_h)^T J_{2n} \Phi'_h = J_{2n}$. Note that the implicit midpoint rule is a symmetric scheme, i.e., $\Phi^*_h = \Phi_h$. This symmetry implies that the method is of an even order [51], therefore is at least of order 2. □

3 Model Order Reduction

Mathematical simulation is increasingly important in engineering, science, and related domains, thanks to substantial advances in computational sciences and the rapid growth in computational capacity during the past decades. Numerical evaluation of partial differential equations (PDEs) lies at the core of these disciplines which includes design, optimization, and prediction of inputs and outputs of interest. However, the need for accuracy, the complexity of multi-physics applications, and inefficiencies in evaluating multi-query systems makes conventional approaches for solving large systems of partial differential equations impractical.

To cope with these limitations, *reduced order modelling* (ROM), apposed to *full-order* or *high-fidelity* modelling, has been an area of active research for the past decade. These methods eliminate the redundant physical or computational complexities of the full-order problem to construct a low dimensional reduced-order system. This approximation in return significantly accelerates the evaluation of the system of PDEs. Reduced basis (RB) methods are among the most successful ROMs and are used throughout academia and industry. RB methods seek a low dimensional reduced subspace that accurately represents the full-order solution. Confining the system to this subspace, through a projection, allows to accelerate the evaluation of the system.

In this chapter we summarize the fundamentals of model order reduction (MOR) and especially RB methods. We present various conventional techniques and algorithms for linear and nonlinear problems. Since time, as a parameter, is particularly important in the context of Hamiltonian systems, we will develop this chapter with an emphasis on time-dependent problems.

3.1 Solution Manifold and Reduced Basis Methods

We consider parametric dynamical systems of the type

$$\begin{cases} \dfrac{d}{dt}u(t;\mu) = f(t,u;\mu), \\ u(0;\mu) = u_0(\mu). \end{cases} \tag{3.1}$$

Here $u, u_0 \in \mathbb{R}^n$, $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{P} \to \mathbb{R}^n$ is a Lipschitz linear or nonlinear function, and $\mu \in \mathbb{P}$, where $\mathbb{P}$ is a compact subset of $\mathbb{R}^d$. It is well known that for a fixed μ, (3.1) has a unique solution if f is continuous with a continuous derivative [92]. Note that for parametric PDEs, we may use the method of lines [33] to obtain a dynamical system of the form (3.1).

To obtain a numerical solution to (3.1) for a fixed μ, we may use some time integration method, e.g. the Runge-Kutta methods [33]. This provides an approximate solution $\tilde{u}(t_i) \approx u(t_i)$ for time instances $i = 1, \ldots, N_t$. Throughout this chapter we assume that $\tilde{u}$ can be obtained arbitrary close to u and, by abuse of the notation, we may omit the overscript "~". In the MOR literature, u is often referred to as the *full-order* or the *high-fidelity* solution [52, 86].

Definition 3.1. *The solution manifold is a set of all solutions to* (3.1) *under the variation of the parameter vector μ, i.e.*

$$\mathcal{M}_u = \{u(t;\mu) | t \in [0,T], \mu \in \mathbb{P}\}. \tag{3.2}$$

Note that the solution manifold may not be smooth. A main assumption in an RB method is that $\mathcal{M}_u$ has a low dimensional representation. This allows us to chose a small number of basis vectors $E_k = \{w_1, \ldots, w_k\}$, with $k \ll n$, where $\mathcal{W}_k = \text{span}(E_k)$ represents $\mathcal{M}_u$ with a small error. E_k is often referred to as the *reduced basis*. To understand when a low dimensional reduced basis exists and to quantify the error in the approximation, we need to introduce the notion of the *Kolmogorov n-width* [63, 83].

Definition 3.2. *Let $\mathcal{W}$ be a subset of a Banach space $\mathcal{X}$. The distance of a point $x \in \mathcal{X}$ from $\mathcal{W}$ is given by*

$$\textit{dist}(x, \mathcal{W}) := \inf_{w \in \mathcal{W}} \|x - w\|. \tag{3.3}$$

where $\|\cdot\|$ is the norm defined on $\mathcal{X}$.

We can look at $\text{dist}(x, \mathcal{W})$ as a measure on how well we can approximate x with elements in $\mathcal{W}$.

Definition 3.3. *Let $\mathcal{S}$ be a compact subset of a Banach space $\mathcal{X}$. The Kolmogorov n-width of S is defined as*

$$d_n(\mathcal{S}) = \inf_{\mathcal{W}_n} \sup_{s\in\mathcal{S}} \textit{dist}(s, \mathcal{W}_n), \tag{3.4}$$

where the infimum is carried over all possible linear subspaces $\mathcal{W}_n$ of dimension n.

Therefore, the n-width measures how well $\mathcal{S}$ can be approximated by a linear subspace of dimension n. Note that when $\mathcal{X}$ is also equipped with an inner product operator $< \cdot, \cdot >: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, such that $\| \cdot \| = \sqrt{< \cdot, \cdot >}$, then dist$(x, \mathcal{W}) = \|x - P_{<,>,\mathcal{W}}(x)\|$, where $P_{<,>,\mathcal{W}}$ is the projection operator with respect to $< \cdot, \cdot >$ onto $\mathcal{W}$. In this case dist$(x, \mathcal{W})$ is often referred to as the *projection error*.

RB method seeks to approximate $\mathcal{M}_u$ with a low dimensional subspace $\mathcal{W}_k$, making it natural to use the n-width terminology. To achieve an accurate RB approximation we truncate the sequence $d_1(\mathcal{M}_u), d_2(\mathcal{M}_u), \ldots, d_n(\mathcal{M}_u)$ such that the truncation error is controlled, i.e.

$$\frac{\sum_{i=1}^{k} d_i(\mathcal{M}_u)}{\sum_{i=1}^{n} d_i(\mathcal{M}_u)} < \delta, \tag{3.5}$$

for some small tolerance δ. Therefore, it is desirable that the above sequence has a rapid decay, in which case $\mathcal{M}_u$ is referred to as *reducible*. In general, the dimension k must be chosen small enough to enable computational gain. Once the subspace $\mathcal{W}_k$ is chosen, we can construct the projection operator $P_{<,>,\mathcal{W}_k}$ to write the reduced-order system

$$\begin{cases} \dfrac{d}{dt} P_{<,>,\mathcal{W}_k}(u(t;\mu)) = P_{<,>,\mathcal{W}_k}(f(t,u;\mu)), \\ \quad P_{<,>,\mathcal{W}_k}(u(0;\mu)) = P_{<,>,\mathcal{W}_k}(u_0(\mu)). \end{cases} \tag{3.6}$$

Note that in this thesis, we assume that the projection operator $P_{<,>,\mathcal{W}_k}$ (and subsequently the reduced space $\mathcal{W}_k$) is not time dependent[1]. Therefore, we may commute the projection operator with the time derivation operator. Given that $k \ll n$, (3.6) has a lower order than (3.1). However, reducibility of $\mathcal{M}_u$ does not say anything about the stability of (3.6). As the matter of fact, (3.6) could be unstable even if (3.1) is a stable dynamical system [81, 2]. In Chapters 4 to 6 we discuss how we can enhance the stability of (3.6) given that (3.1) is a Hamiltonian system.

In the following we summarize numerical methods for choosing the dimension k, finding the reduced space $\mathcal{W}_k$, constructing the projection operator $P_{<,>,\mathcal{W}_k}$ and, finally, efficient ways to construct and integrate the reduced system (3.6).

[1] We refer the reader to [78, 38] for dynamically orthogonal reduced basis method: an RB method with a time varying basis.

3.2 Proper Orthogonal Decomposition

To numerically recover $\mathcal{W}_k$, we first discretize the solution manifold $\mathcal{M}_u$ into a point cloud $\mathcal{M}_u^\Delta$ defined as

$$\mathcal{M}_u^\Delta := \{u(t_i;\mu) | 1 \leq i \leq N_t \text{ and } \mu \in \mathbb{P}^\Delta\}, \tag{3.7}$$

where $\mathbb{P}^\Delta = \{\mu_1, \ldots, \mu_{N_\mathbb{P}}\}$ is a finite set, representing $\mathbb{P}$. Note that the choice of $\mathbb{P}^\Delta$ is generally not trivial and is often problem dependent. We refer the reader to [86] for more information on discretizing the parameter space.

Each member of $\mathcal{M}_u^\Delta$ is a vector in $\mathbb{R}^n$ and is commonly referred to as a *snapshot*. Suppose that we can find k basis vectors $w_1, \ldots, w_k \in \mathbb{R}^n$, orthonormal with respect to some inner product operator $< \cdot, \cdot >$, and with a space $\mathcal{W}_k$ which approximately represents the span($\mathcal{M}_u^\Delta$). As discussed in Section 3.1, the projection error of a member of $\mathcal{M}_u^\Delta$ by an element in $\mathcal{W}_k$ is given by

$$e_{\mathcal{W}_k}(s) := \|s - P_{<,>,\mathcal{W}_k}(s)\|, \quad s \in \mathcal{M}_u^\Delta, \tag{3.8}$$

where $\|\cdot\|$ is the norm associated with $< \cdot, \cdot >$ and $P_{<,>,\mathcal{W}_k}$ is the orthogonal projection operator given by

$$P_{<,>,\mathcal{W}_k}(s) = \sum_{i=1}^{k} < s, w_i > w_i. \tag{3.9}$$

The *proper orthogonal decomposition* method then identifies $\mathcal{W}_k$ (for a fixed k) that minimizes the collective projection error, corresponding to the minimization problem

$$\begin{aligned} &\underset{\mathcal{W}_k}{\text{minimize}} \quad \sum_{s \in \mathcal{M}_u^\Delta} \|s - P_{<,>,\mathcal{W}_k}(s)\|^2, \\ &\text{subject to} \quad < w_i, w_j >= \delta_{i,j}, \quad 1 \leq i, j \leq k. \end{aligned} \tag{3.10}$$

This formulation is comparable with the discrete version of the Kolmogorov n-width.

3.2.1 Euclidean Inner Product

When $< \cdot, \cdot >$ is the classical Euclidean inner product, i.e. $< a, b >= a^T b$ for $a, b \in \mathbb{R}^n$, we can rewrite the projection operator in (3.9) as

$$P_{I,\mathcal{W}_k}(s) = W_k W_k^T s. \tag{3.11}$$

Here $W_k = [w_i]_{i=1}^k$ is the matrix containing the basis vectors of $\mathcal{W}_k$ (Note that we used the subscript I to indicate the Euclidean inner product in $P_{I,\mathcal{W}_k}$. To avoid confusion, we may also use this subscript for $< \cdot, \cdot >_I$). Furthermore, the constraints in minimization

(3.10) simplify to $W_k^T W_k = I_k$. Thus, (3.10) becomes

$$\begin{aligned} &\underset{W_k \in \mathbb{R}^{n\times k}}{\text{minimize}} \quad \sum_{s\in\mathcal{M}_u^\Delta} \|s - W_k W_k^T s\|_2^2, \\ &\text{subject to} \quad W_k^T W_k = I_k. \end{aligned} \tag{3.12}$$

Here $\|\cdot\|_2$ is the Euclidean norm. Finally, if we collect all the snapshots into the *snapshot matrix*

$$S = [u(t_i;\mu_j)], \quad 1 \le i \le N_t, \quad 1 \le j \le N_{\mathbb{P}}, \tag{3.13}$$

we can use basic results in linear algebra [104] to reformulate (3.10) as

$$\begin{aligned} &\underset{W_k \in \mathbb{R}^{n\times k}}{\text{minimize}} \quad \|S - W_k W_k^T S\|_F, \\ &\text{subject to} \quad W_k^T W_k = I_k. \end{aligned} \tag{3.14}$$

Here $\|\cdot\|_F$ denotes the Frobenius norm [104].This minimization is nonlinear and generally non-convex. However, a remarkable result in numerical analysis relates the solution to this minimization problem with an eigenvalue problem on S. We summarize this in the following theorem and refer the reader to [68] for the proof.

Theorem 3.1. *(Eckart-Young-Mirsky-Schmidt) Suppose that $D \in \mathbb{R}^{m\times n}$ $(m < n)$ has the singular value decomposition (SVD) [68], $D = U\Sigma V^T$. Consider the partitioning for U, Σ and V as*

$$U = [U_1 U_2], \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \quad V = [V_1 V_2], \tag{3.15}$$

where $U_1 \in \mathbb{R}^{n\times r}$, $U_2 \in \mathbb{R}^{n\times(n-r)}$, $\Sigma_1 \in \mathbb{R}^{r\times r}$, $\Sigma_2 \in \mathbb{R}^{(n-r)\times(n-r)}$, $V_1 \in \mathbb{R}^{n\times r}$ and $V_2 \in \mathbb{R}^{n\times(n-r)}$. Then the rank r matrix, resulting from the truncation of the SVD decomposition

$$\tilde{D} = U_1 \Sigma_1 V_1^T, \tag{3.16}$$

solves the minimization problem

$$\begin{aligned} &\underset{M\in\mathbb{R}^{m\times n}}{\textit{minimize}} \quad \|D - M\|_F, \\ &\textit{subject to} \quad \textit{rank}(M) = r. \end{aligned} \tag{3.17}$$

Furthermore, $\|D - \tilde{D}\|_F = \sum_{i=r+1}^{m} \sigma_i$, where σ_i is the i-th singular value of D.

With this, we immediately find the solution to (3.14).

Theorem 3.2. *[4] Let $S = U\Sigma V^T$ be the SVD decomposition of S with $U = [u_i]_{i=1}^n$. Then*

$W = U_1$ is the solution to the minimization problem (3.14) *where $\tilde{S} = U_1\Sigma_1 V_1^T$ is the rank k approximation of S in Theorem 3.1.*

Proof. Let $\tilde{S} = U_1\Sigma_1 V_1^T$ be the matrix that minimizes $\|S - \tilde{S}\|_F$ in Theorem 3.1 and let $\tilde{\Sigma}$ be defined as

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.18}$$

It is easily verified that $\tilde{S} = U_1\Sigma_1 V_1^T = U\tilde{\Sigma}V^T$. This yields

$$\tilde{S} = U\tilde{\Sigma}V^T = U\tilde{\Sigma}\Sigma^{-1}U^T S = U_1 U_1^T S. \tag{3.19}$$

Therefore, minimizing $\|S - \tilde{S}\|_F$ for a rank k matrix $\tilde{S}$ is equivalent to minimizing $\|S - U_1 U_1^T S\|_F$ for all U_1 such that $U_1^T U_1 = I_k$. □

As the minimization problem (3.14) is closely related to the Kolmogorov n-width of $\mathcal{M}_u$, Theorem 3.1 suggests that the decay of the singular values of S is an indicator of the decay of the Kolmogorov n-width of $\mathcal{M}_u$. This is a numerical approach to understand the reducibility of $\mathcal{M}_u$. Algorithm 3.1 summarizes POD to generate an orthonormal basis with respect to the Euclidean inner product.

Algorithm 3.1 POD for constructing an orthonormal reduced basis with respect to the Euclidean inner product

Input: snapshot matrix S, tolerance value δ.

1: compute the SVD decomposition $S = U\Sigma V^T$.
2: pick k as the largest number such that

$$\frac{\sum_{i=k+1}^{n}\sigma_i}{\sum_{i=1}^{n}\sigma_i} < \delta.$$

3: define $W_k = [u_i]_{i=1}^k$.

Output: reduced basis W_k.

3.2.2 Non-Euclidean Inner Product

Now suppose that $<,>$ is a non-Euclidean inner product. One can associate such an inner product with a symmetric and positive definite matrix X (and denote $<,>_X$) such that $< a, b >_X = a^T X b$ for all $a, b \in \mathbb{R}^n$. Given an orthonormal basis matrix $W_k = [w_i]_{i=1}^k$ with respect to this inner product, it is easy to verify that the orthogonal

projection onto $\mathcal{W}_k = \text{col span}(W_k)$ is given by

$$P_{X,\mathcal{W}_k}(s) = W_k W_k^T X s, \quad s \in \mathbb{R}^n. \tag{3.20}$$

With this, the minimization problem (3.10) becomes

$$\begin{aligned} &\underset{W_k \in \mathbb{R}^{n\times k}}{\text{minimize}} \quad \sum_{s\in\mathcal{M}_u^\Delta} \|s - W_k W_k^T X s\|_X^2, \\ &\text{subject to} \quad W_k^T X W_k = I_k. \end{aligned} \tag{3.21}$$

Here $\|\cdot\|_X = \sqrt{<\cdot,\cdot>_X}$ and the constraint arises from $< w_i, w_j >_X = w_i^T X w_j = \delta_{ij}$ for $i = 1, \ldots, k$. It follows that

$$\begin{aligned} \sum_{s\in\mathcal{M}_u^\Delta} \|s - W_k W_k^T X s\|_X^2 &= \sum_{s\in\mathcal{M}_u^\Delta} \|X^{1/2} s - X^{1/2} W_k W_k^T X s\|_2^2 \\ &= \sum_{s\in\mathcal{M}_u^\Delta} \|X^{1/2} s - X^{1/2} W_k W_k^T X^{1/2} X^{1/2} s\|_2^2 \\ &= \|X^{1/2} S - X^{1/2} W_k W_k^T X^{1/2} X^{1/2} S\|_F^2. \end{aligned} \tag{3.22}$$

Here $X^{1/2}$ is the matrix square root of X. Now if we define $\bar{S} = X^{1/2} S$ and $\bar{W}_k = X^{1/2} W_k$, we can rewrite (3.21) as

$$\begin{aligned} &\underset{\bar{W}_k \in \mathbb{R}^{n\times k}}{\text{minimize}} \quad \|\bar{S} - \bar{W}_k \bar{W}_k^T \bar{S}\|_F, \\ &\text{subject to} \quad \bar{W}_k^T \bar{W}_k = I_k. \end{aligned} \tag{3.23}$$

Following Theorems 3.1 and 3.2, the solution $\bar{W}_k$ to this minimization problem is the rank k approximation of $\bar{S}$. We can then recover W_k form $W_k = X^{-1/2} \bar{W}_k$. Constructing a POD basis with respect to a non-Euclidean inner product is presented in Algorithm 3.2.

For large scale problems, the computation of the square root of X may be computationally demanding. Let $\bar{S} = U\Sigma V$ be the SVD decomposition of $\bar{S}$ with $\{u_i\}_{i=1}^n$, $\{v_i\}_{i=1}^n$, and $\{\sigma_i\}_{i=1}^n$ the left singular vectors, the right singular vectors, and the singular values, respectively. We have

$$S^T X S v_i = \bar{S}^T \bar{S} v_i = \sigma_i \bar{S}^T u_i^T = \sigma_i^2 v_i. \tag{3.24}$$

Here we used the properties of an SVD decomposition [104]. This suggests that $\{\sigma_i^2\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ are the eigenvalues and the eigenvectors of $S^T X S$, respectively. The matrix $G := S^T X S$ is commonly referred to as the *Gramian matrix*. To obtain the POD basis, we can then write

$$w_i = X^{-1/2} u_i = \sigma_i^{-1} X^{-1/2} \bar{S} v_i = \sigma_i^{-1} S v_i. \tag{3.25}$$

Algorithm 3.2 POD for constructing an orthonormal reduced basis with respect to a non-Euclidean inner product

Input: snapshot matrix S, weight matrix X, and tolerance value δ.

1: compute $\bar{S} = X^{1/2}S$.
2: compute the SVD decomposition $\bar{S} = U\Sigma V^T$.
3: pick k as the largest number such that

$$\frac{\sum_{i=k+1}^{n} \sigma_i}{\sum_{i=1}^{n} \sigma_i} < \delta.$$

4: define $\bar{W}_k = [u_i]_{i=1}^{k}$.
5: compute $W_k = X^{-1/2}\bar{W}_k$.

Output: reduced basis W_k.

Thus, the computation of $X^{1/2}$ can be avoided. We summarize the computationally efficient way to find a POD basis with respect to a non-Euclidean inner product in Algorithm 3.3.

Algorithm 3.3 POD for constructing an orthonormal reduced basis with respect to a non-Euclidean inner product

Input: snapshot matrix S, weight matrix X.

1: compute the Gramian matrix $G = S^T XS$.
2: solve the eigenvalue problem $Gv_i = \sigma_i^2 v_i$.
3: compute $w_i = \sigma_i^{-1} S v_i$.
4: define basis $W_k = [w_i]_{i=1}^{k}$.

Output: reduced basis W_k.

3.3 The Greedy Basis Generation

For the purpose the of most efficient computation, it is important to separate expensive high-dimensional quantities from the cheaper low-dimensional ones. This Separation of quantities, according to their computational cost, is referred to as the *offline/online* decomposition [86]. We tolerate some amount of computational complexity in the offline phase to achieve substantial computation acceleration during the online phase. In the context of RB methods, one seeks to restrict computations regarding the high-fidelity solution to the offline phase. Subsequently, we can expect fast computations by restricting all computations to the reduced space during the online phase.

Assembling the snapshot matrix S (3.13) requires the evaluation of the high-fidelity solution for possibly a large sample set of the parameter space $\mathbb{P}$. Moreover, performing an SVD decomposition on S can be computationally demanding and often impractical

during the offline phase. A greedy basis generation is an iterative process in which basis vectors are identified and added, one at a time, to improve the overall accuracy of the reduced basis. As the high-fidelity solution is only evaluated once per iteration, the assembly and the SVD decomposition of S is avoided, saving considerable computational effort in the offline phase. Note that Theorem 3.1 indicates that the SVD provides the best possible basis of a given size k in L^2. The greedy approach, on the other hand, provides an optimal basis in L^∞. However, the two methods often result in a basis of a comparable size and accuracy [86].

Since a key step in a greedy method is the identification of the best possible candidate for a basis vector, the availability of an error indicator is essential. Let $W_k = [w_i]_{i=1}^k$ be an orthonormal reduced basis with $\mathcal{W}_k$ as the reduced space spanned by the column vectors of W_k. Inspired by the Kolmogorov n-width, we use the projection error to identify the snapshot s that is worst approximated by a member of $\mathcal{W}_k$:

$$s^* := \arg\max_{s\in S} \operatorname{dist}(s, \mathcal{W}_k) = \arg\max_{s\in S} \|s - P_{X,\mathcal{W}_k}(s)\|_X. \tag{3.26}$$

Here, s^* is then a candidate for the next basis vector. Let w_{k+1} be the vector obtained by orthonormalizing s^* with respect to W_k. The next basis matrix is then defined as

$$W_{k+1} = [w_i]_{i=1}^{k+1}, \quad \mathcal{W}_{k+1} = \operatorname{col\,span}(W_{k+1}). \tag{3.27}$$

Note that the choice of an orthonormalization process is important when constructing a reduced basis with a low condition number. Therefore, a backward stable method is preferred. The process discussed above is referred to as the *strong greedy method* [86]. The following theorem from approximation theory shows that the convergence rate of the strong greedy method, is as fast as the decay of the Kolmogorov n-width.

Theorem 3.3. *[16] Let $S \subset \mathbb{R}^n$ have an exponentially small Kolmogorov n-width $d_n(S) \leq c \cdot \exp(-\alpha n)$ with $\alpha > \log 2$. Then there exists $\beta > 0$ such that the subspace $\mathcal{W}_k$ constructed by the strong greedy process is exponentially accurate in the sense*

$$\|s - P_{X,\mathcal{W}_k}(s)\|_X \leq C \cdot \exp(-\beta k), \quad s \in S. \tag{3.28}$$

The proof for this theorem is lengthy but straight forward. Orthonormality of $\mathcal{W}_k$ is exploited in the proof. Therefore, for cases where a non-orthonormal reduced basis is considered, additional constraints must be checked to guarantee convergence of the method. It is worth mentioning that the inequality (3.28) has been improved in [19] under further assumptions on the reducibility of S.

The maximization problem (3.26) still contains computational inefficiencies, since the high-fidelity solution s needs to be evaluated over the entire parameter space. The *weak greedy method* avoids this by finding a surrogate function $\eta : \mathbb{P} \to \mathbb{R}$ that

approximates (3.26) at a low cost. This restricts the search for the new basis vector to only the parameter space, rather than the high-fidelity space. Let $M_k = \{\mu_i\}_{i=1}^k$ be a set of chosen parameters and $E_k = \{w_i\}_{i=1}^{n_k}$ be the set of basis vectors with $\mathcal{W}_k$ as the corresponding reduced space. Using η, the next parameter can be chosen as

$$\mu_{k+1} := \arg\max_{\mu \in \mathbb{P}} \eta(\mu). \tag{3.29}$$

We may then compute the snapshots $S_{t,\mu_{k+1}} = [u^*(t_i)]_{i=1}^{N_t}$, where $u^*(t_i) = u(t_i, \mu_{k+1})$. We eliminate the common subspace between the spaces of $S_{t,\mu_{k+1}}$ and $\mathcal{W}_k$ by computing the errors

$$e_{\mu_{k+1}} = [u^*(t_i) - P_{X,\mathcal{W}_k}(u^*(t_i))]_{i=1}^{N_t}. \tag{3.30}$$

For a given tolerance δ, we add the truncated POD basis vectors of $e_{\mu_{k+1}}$, e.g. vectors $\{w_i^{\mu_{k+1}}\}_{i=1}^{n_{m+1}}$, to the previously computed basis vectors

$$E_{k+1} = E_k \cup \{w_i^{\mu_{k+1}}\}_{i=1}^{m_{k+1}}, \quad \mathcal{W}_{k+1} = \text{span}(E_{k+1}), \quad n_{k+1} = n_k + m_{k+1}, \tag{3.31}$$

and we denote by W_{k+1} the corresponding basis matrix. This approach is referred to as the *POD-greedy* method [49]. Given a proper error indicator function η, the *POD-greedy* method can also provide an exponentially accurate reduced space [49, 47]. We summarize the POD-greedy method in Algorithm 3.4 and refer the reader to [52, 47] for further details.

Algorithm 3.4 the POD-greedy for extending a reduced basis

Input: parameter space $\mathbb{P}$, reduced basis W_k of size n_k, truncation tolerance δ.

1: find $\mu^* := \arg\max_{\mu \in \mathbb{P}} \eta(\mu)$.
2: compute temporal snapshots S_{t,μ^*}.
3: compute the error vectors E_{μ^*}.
4: $W_{k+1} \leftarrow W_k \cup POD(E_{\mu^*}, \delta)$.
5: $n_{k+1} \leftarrow n_k + m_{k+1}$.

Output: reduced basis W_{k+1}.

3.4 The Galerkin and the Petrov-Galerkin Projection

In Sections 3.2 and 3.3 we outlined computational methods to construct a reduced basis W_k, and subsequently, a reduced space $\mathcal{W}_k$. In this section we elaborate on how to use a reduced basis to construct a reduced system, as in (3.6).

Let W_k be an orthonormal basis for a reduced space $\mathcal{W}_k$. We assume $u(t;\mu)$, the solution to (3.1), can be well approximate in this basis as $u \approx \tilde{u} = W_k v$, where $v \in \mathbb{R}^k$ is the

vector of coefficients. Substituting this in (3.1) results in

$$W_k \frac{d}{dt} v(t;\mu) = f(t, W_k v; \mu) + r(t, u; \mu), \tag{3.32}$$

and r is the error in this approximation. Now, we can use the properties of the projection operator $P_{I,\mathcal{W}_k} = W_k W_k^T$ to eliminate W_k from the left hand side. The *Galerkin* projection requires r to be orthogonal to W_k and results in the reduced system

$$\begin{cases} \frac{d}{dt} v(t;\mu) = W_k^T f(t, W_k v; \mu), \\ v(0;\mu) = W_k^T u_0(\mu). \end{cases} \tag{3.33}$$

Here we used the fact that $W_k^T W_k = I_k$. We may instead use a non-Euclidean inner product, with the projection operator $P_{X,\mathcal{W}_k} = W_k W_k^T X$, to obtain

$$\begin{cases} \frac{d}{dt} v(t;\mu) = W_k^T X f(t, W_k v; \mu), \\ v(0;\mu) = W_k^T X u_0(\mu). \end{cases} \tag{3.34}$$

The *Petrov-Galerkin* projection, on the other hand, requires r in (3.32) to be orthogonal to some k-dimensional linear subspace $\mathcal{U}$. Given U as the basis matrix for this subspace, and requiring $U^T W_k$ to be invertible, a projection operator that projects the elements of the high-fidelity space onto $\mathcal{W}_k$ orthogonal to $\mathcal{U}$ can be constructed as $\Pi = W_k (U^T W_k)^{-1} U$. This results in the reduced system

$$\begin{cases} \frac{d}{dt} v(t;\mu) = (U^T W_k)^{-1} f(t, W_k v; \mu), \\ v(0;\mu) = (U^T W_k)^{-1} u_0(\mu). \end{cases} \tag{3.35}$$

Although reduced systems in (3.33), (3.34) and (3.35) are of a lower order as compared to the high-fidelity system, the evaluation of $f(t, W_k v; \mu)$ should be performed in the high-fidelity space for a general f. In the following section we discuss how this can be avoided.

3.5 Efficient Evaluation of the Non-Linear Terms

In this section we discuss the efficiency of evaluating nonlinear terms in the context of projection based reduced models. Suppose that the right hand side in (3.1) is of the form $f(t, u; \mu) = L(\mu)u + g(t, u; \mu)$, where L reflects the linear part, and g is a nonlinear function. Now assume that a k-dimensional reduced basis $W \in \mathbb{R}^{n \times k}$ is provided. The

reduced system using a Petrov-Galerkin projection takes the form

$$\frac{d}{dt}v = \underbrace{U^T L W}_{\tilde{L}} v + \underbrace{U^T g(t, Wv; \mu)}_{h(t,v;\mu)}. \tag{3.36}$$

Here, $\tilde{L}$ is a $k \times k$ matrix which can be computed in the offline phase. However, the evaluation of $h(t, v; \mu)$ has a complexity that depends on n, the size of the original system. Suppose that the evaluation of g with n components has the complexity $\alpha(n)$, for some function α. Then the complexity of evaluating h is $\mathcal{O}(\alpha(n) + 4nk)$ which consists of 2 matrix-vector operations and the evaluation of the nonlinear function, i.e. the evaluation of the nonlinear terms can be as expensive as solving the original system.

To overcome this bottleneck we take a reduced basis approach [27, 8]. Assume that the manifold $\mathcal{M}_g = \{g(t, u; \mu) | t \in [0, T], u \in \mathbb{R}^n, \mu \in \mathbb{P}\}$ is low dimensional and that g can be approximated by a linear subspace of dimension $m \ll n$, spanned by the basis $\{y_i\}_{i=1}^m$, i.e.

$$g \approx Yc. \tag{3.37}$$

Here $Y \in \mathbb{R}^{n \times m}$ contains basis vectors y_i and $c \in \mathbb{R}^m$ is the vector of coefficients. Now suppose $p_1, \ldots, p_m$ are m indices from $\{1, \ldots, n\}$ and define an $n \times m$ real matrix

$$P := [e_{p_1}, \ldots, e_{p_m}], \tag{3.38}$$

where e_{p_i} is the p_i-th column of the identity matrix I_n. Multiplying P^T with g selects components $p_1, \ldots, p_m$ of g. If we assume that $P^T U$ is non-singular, the coefficient vector c can be uniquely determined from

$$P^T \tilde{g} = (P^T Y)c. \tag{3.39}$$

Here, the overscript "~" emphasises that $\tilde{g}$ is an approximation to g. Finally we have

$$g(t, u; \mu) \approx Y(\mu)c(t, u) = Y(P^T Y)^{-1} P^T g(t, u; \mu), \tag{3.40}$$

which is referred to as the *Empirical Interpolation* (EIM) approximation [8]. Applying EIM to the reduced system (3.36) yields

$$\frac{d}{dt}u = \tilde{L}u + U^T Y(P^T Y)^{-1} P^T g(t, u; \mu). \tag{3.41}$$

Note that the matrix $U^T Y(P^T Y)^{-1}$ can be computed offline and since g is evaluated only at m of its components, the evaluation of the nonlinear term in (3.41) does not depend on n.

To obtain the projection basis U, the POD can be applied to the matrix S_g that contains the snapshots of the nonlinear term g

$$S_g = [g(t_i, u; \mu_j)], \quad 1 \leq i \leq N_t, \quad 1 \leq j \leq N_{\mathbb{P}}. \tag{3.42}$$

Note that there is no additional cost associated with computing these snapshots, since they are generated when computing the trajectory snapshot matrix S.

The interpolating indices $p_1, \ldots, p_m$ can be constructed as follows. Given the projection basis $Y = \{y_1, \ldots, y_m\}$, the first interpolation index p_1 is chosen according to the component of u_1 with the largest magnitude. The rest of the interpolation indices, $p_2, \ldots, p_m$ correspond to the component of the largest magnitude of the residual vector $\mathbf{r} = y_l - Yc$. It is shown in [27] that if the residual vector is a nonzero vector in each iteration then $P^T U$ is non-singular and thus the reduced system (3.41) is well defined. The application of the POD for generating a basis for the nonlinear term together with the greedy selection of interpolating indices is referred to as the *Discrete Empirical Interpolation Method* (DEIM). We summarized the process of selecting interpolating indices for DEIM in Algorithm 3.5.

Algorithm 3.5 Discrete Empirical Interpolation Method

Input: Basis vectors $\{u_1, \ldots, u_m\} \subset \mathbb{R}^n$

pick p_1 to be the index of the largest component of u_1.
$U \leftarrow [u_1]$.
$P \leftarrow [p_1]$.
for $i \leftarrow 2$ **to** m
    solve $(P^T U)\mathbf{c} = P^T u_i$ for $\mathbf{c}$.
    $\mathbf{r} \leftarrow u_i - U\mathbf{c}$.
    pick p_i to be the index of the largest component of $\mathbf{r}$.
    $U \leftarrow [u_1, \ldots, u_i]$.
    $P \leftarrow [p_1, \ldots, p_i]$.
end for

Output: Interpolating indices $\{p_1, \ldots, p_m\}$.

The numerical integration of (3.41) may involve the computation of the Jacobian of the nonlinear function $g(t, u; \mu)$ with respect to the reduced state variable $\mathbf{v}$

$$\mathcal{J}_v(g) = U^T \mathcal{J}_u(g) W, \tag{3.43}$$

where $\mathcal{J}_\alpha(g)$ is the Jacobian matrix of g with respect to the variable $\alpha = u, v$. The complexity of (3.43) is $\mathcal{O}(\alpha(n) + 2n^2k + 2nk^2 + 2nk)$, comprising several matrix-vector multiplications and an evaluation of the Jacobian which depends on the size of the original system. Approximating the Jacobian in (3.43) is usually both problem and discretization dependent. Often the nonlinear function g is evaluated component-wise

i.e.

$$\mathbf{g}(u) = \begin{pmatrix} g_1(u_1, \ldots, u_n) \\ g_2(u_1, \ldots, u_n) \\ \vdots \\ g_n(u_1, \ldots, u_n) \end{pmatrix} = \begin{pmatrix} g_1(u_1) \\ g_2(u_2) \\ \vdots \\ g_n(u_n) \end{pmatrix}. \tag{3.44}$$

In such cases the interpolating index matrix P and the nonlinear function g commute, i.e.,

$$h(v) \approx U^T Y (P^T Y)^{-1} P^T g(Wv) = U^T Y (P^T Y)^{-1} \mathring{g}(P^T Wv) \tag{3.45}$$

Here, $\mathring{g}$ indicates that the elements of g, that are not in the index set P, are omitted. If we now take the Jacobian of the approximate function we recover

$$\mathcal{J}_v(g) = U^T Y (P^T Y)^{-1} \mathcal{J}_u(\mathring{g}(P^T Wv)) P^T W. \tag{3.46}$$

The matrices $U^T Y (P^T Y)^{-1} \in \mathbb{R}^{k \times m}$ and $P^T W \in \mathbb{R}^{m \times k}$ can be computed offline and the Jacobian is evaluated only for $m \times m$ components. Hence the overall complexity of computing the Jacobian is now independent of n. We refer the reader to [27, 8, 86, 52] for more detail.

4 Symplectic Model Order Reduction

Parameterized partial differential equations often arise as a model for many problems in engineering and the applied sciences. While the need for more accuracy has led to the development of exceedingly complex models, the limitations in computational cost and storage often make direct approaches impractical. Hence, we must seek alternative methods, such as those introduced in Chapter 3, that allow us to approximate the desired output under variation of the input parameters while keeping the computational costs to a minimum.

As we discussed in Chapter 3, SVD based model reduction method, such as POD, often require the exploration of the entire parameter space [64, 5, 87, 55, 56, 57, 82]. This leads to a very expensive and often impractical offline stage when dealing with multi-dimensional parameter domains. On the other hand, sampling techniques, usually of a greedy nature, search through the parameter space selectively, guided by an error estimate to certify the accuracy of the basis. This approach, accompanied with an efficient sampling procedure, balances the cost of computation with the overall accuracy of the reduced-basis [29, 91, 52].

Besides computational complexity, another aspect of reduced order modeling is the preservation of structure and, in particular, the stability of the original model. In general, reduced order models of the type discussed in Chapter 3 do not guarantee that such properties are preserved [85].

In the context of Hamiltonian and Lagrangian systems, recent work suggests modifications of POD to preserve some geometric structures. Lall et al. [65] and Carlberg et al. [24] suggest that the reduced-order system should be identified by a Lagrangian function on a low-dimensional configuration space. In this way, the geometric structure of the original system is inherited by the reduced system. Model reduction for port-Hamiltonian systems can be found in the works of Beattie et al. [26], Polyuga et al. [84] and references therein. These works construct a reduced port-Hamiltonian system using Krylov or POD methods that inherit the passivity and stability of the original

system. For Hamiltonian systems, Peng et al. [81], using a symplectic transformation, constructs a reduced Hamiltonian as an approximation to the Hamiltonian of the original system. As a result, the reduced system preserves the symplectic structure. Although these methods preserve the geometric structure, they use a POD-like approach for constructing the reduced basis. If the numerical evaluation of the original model is computationally demanding, performing POD can be excessively expensive [86].

In this chapter, we present a greedy approach for the construction of a reduced system that preserves the geometric structure of Hamiltonian systems. This technique results in a reduced Hamiltonian system that mimics the symplectic properties of the original system and preserves the Hamiltonian structure and its stability over the course of time. On the other hand, since time integration of the original system is only required once per iteration, the proposed method saves substantial computational cost during the offline stage as compared to alternative POD-like techniques. It is well known that structured matrices, e.g. symplectic matrices, generally are not well-conditioned [61]. The greedy update of the symplectic basis presented here yields a ortho-symplectic basis and is therefore a norm bounded. Moreover, we demonstrate that assumptions, natural for the set of all solutions of the original Hamiltonian system under the variation of parameters, lead to exponentially fast convergence of the greedy algorithm. For nonlinear Hamiltonian systems, we show how the basis can be combined with the DEIM [27, 8] to enable a fast evaluation of nonlinear terms while maintaining the symplectic structure.

This chapter is organized as follows. In Section 4.1 the symplectic Galerkin projection to construct a Hamiltonian reduced system is discussed. Sections 4.2 and 4.3 discuss the greedy generation of a symplectic reduced basis as well as other SVD-based symplectic model reduction techniques. Accuracy, stability, and efficiency of the greedy method compared to other SVD-based methods are discussed in Section 4.6. Finally we offer some concluding remarks in Section 4.7.

4.1 Symplectic Galerkin Projection

We now discuss how to modify conventional MOR methods to ensure that the resulting scheme preserves the symplectic structure of the Hamiltonian system.

Let $(\mathcal{Z}, \Omega)$ be a $2n$-dimensional symplectic linear vector space and $H : \mathcal{Z} \to \mathbb{R}$ be a smooth Hamiltonian function. Furthermore, suppose that there is a canonical basis $Z = \{e_i, f_i\}_{i=1}^n$ such that the Hamiltonian equation of evolution takes the form

$$\begin{cases} \dfrac{d}{dt} z = \mathbb{J}_{2n} \nabla_z H(z), \\ z(0) = z_0. \end{cases} \tag{4.1}$$

Here $z = [q_1, \ldots, q_n, p_1, \ldots, p_n]^T \in \mathcal{Z}$ is the state vector (compare to (2.25)). Suppose that the solution manifold $\mathcal{M}_H$ of (4.1) is well approximated by a low dimensional symplectic subspace $(\mathcal{A}, \Omega)$ of dimension $2k$ ($k \ll n$). We can then construct a symplectic basis $E = \{\tilde{e}_i, \tilde{f}_i\}_{i=1}^k$ for $\mathcal{A}$ and assemble the matrix

$$A = [\tilde{e}_1, \ldots, \tilde{e}_k, \tilde{f}_1, \ldots, \tilde{f}_k] \in \mathbb{R}^{2n \times 2k}, \tag{4.2}$$

to approximate the solution to (4.1) as

$$z \approx Ay. \tag{4.3}$$

This constructs a mapping $\phi : \mathcal{A} \to \mathcal{Z}$ given by $\phi(y) = Ay$. Symplecticity of A implies that a symplectic differential form is available on $\mathcal{A}$. One can use the pull-back of Ω along ϕ to construct a symplectic form on $\mathcal{A}$:

$$\tilde{\Omega} = \phi^* \Omega = A^T \mathbb{J}_{2n} A = \mathbb{J}_{2k}. \tag{4.4}$$

Note that in (2.29) a linear symplectic transformation was a square matrix, however, here A is a rectangular matrix. Since $\tilde{\Omega}$ is a symplectic differential form, $(\mathcal{A}, \tilde{\Omega})$ forms a symplectic linear vector space. Furthermore, since the matrix form of $\tilde{\Omega}$ is $\mathbb{J}_{2k}$ we can construct a projection operator (similar to the projection used in the symplectic GS process Corollary 2.8) as

$$P_{\mathcal{A}}^{\text{symp}}(z) = \sum_{i=1}^{k} -\Omega(z, \tilde{f}_i)\tilde{e}_i + \Omega(z, \tilde{e}_i)\tilde{f}_i, \quad z \in \mathcal{Z}. \tag{4.5}$$

It is easily verified that $P_{\mathcal{A}}^{\text{symp}}$ is idempotent, i.e., that $P_{\mathcal{A}}^{\text{symp}} \circ P_{\mathcal{A}}^{\text{symp}} = P_{\mathcal{A}}^{\text{symp}}$. Therefore, this operator is indeed a projection operator onto $\mathcal{A}$ and is known as the *symplectic projection*. In matrix notation, the symplectic projection takes the form

$$P_{\mathcal{A}}^{\text{symp}} = A \mathbb{J}_{2k}^T A^T \mathbb{J}_{2n}. \tag{4.6}$$

The matrix $A^+ := \mathbb{J}_{2k}^T A^T \mathbb{J}_{2n}$ is known as the *symplectic inverse* of A in canonical coordinates. We summarize the properties of A^+ in the following proposition.

Proposition 4.1. *[81] Let $A \in \mathbb{R}^{2n \times 2k}$ be a symplectic matrix in a canonical coordinate system. With this notation the symplectic projection takes the form $P_{\mathcal{A}}^{symp} = AA^+$. The symplectic inverse of A satisfies the following*

(a) $A^+ A = I$.

(b) $\left(\left((A^+)^T\right)^+\right)^T = A$.

(c) $(A^+)^T$ *is a symplectic matrix.*

(d) If A is an ortho-symplectic of the form $A = [e_1, \dots, e_k, \mathbb{J}_{2n}^T e_1, \dots, J_{2n}^T e_k]$, then $(A^+)^T = A$.

Proof. The proof follows immediately from symplecticity of A. □

We can use the symplectic projection operator to project (4.1) onto $\mathcal{A}$. Substituting (4.3) into (4.1) yields

$$A\frac{d}{dt}y = \mathbb{J}_{2n}\nabla_z H(Ay) + r(z). \tag{4.7}$$

We use the chain rule and property (b) in Proposition 4.1 to write

$$\nabla_z H(Ay) = (A^+)^T \nabla_y H(Ay). \tag{4.8}$$

Note that the symplectic inverse is not unique, in general, for a non-square matrix. Furthermore, the symplectic projection yields

$$\frac{d}{dt}y = A^+\mathbb{J}_{2n}(A^+)^T\nabla_y H(Ay) + A^+ r(z). \tag{4.9}$$

Proposition 4.1 (a) ensures that $(A^+)^T$ is a symplectic matrix i.e., $A^+\mathbb{J}_{2n}(A^+)^T = \mathbb{J}_{2k}$. By defining the reduced Hamiltonian $\tilde{H} : \mathcal{A} \to \mathbb{R}$ as $\tilde{H}(y) = H(Ay)$ and assuming that the error vector r is symplectically orthogonal ($\mathbb{J}_{2n}$-orthogonal) to $\mathcal{A}$ we obtain the reduced system

$$\begin{cases} \dfrac{d}{dt}y = \mathbb{J}_{2k}\nabla_y \tilde{H}(y), \\ y_0 = A^+ z_0. \end{cases} \tag{4.10}$$

Equation (4.10) is called the *symplectic Galerkin projection* of (4.1) onto $\mathcal{A}$. The reduced system obtained from the Petrov-Galerkin projection in (3.36) is not a Hamiltonian system and does not guarantee conservation of the symplectic structure, volume of the phase space, or the Hamiltonian. On the other hand, we observe that the reduced system in (4.10) is a Hamiltonian system, and therefore, the symplectic structure will be conserved along integral curves of (4.10). Note that the high-fidelity system and the reduced system are endowed with different Hamiltonians. In the next proposition we show that the error in the Hamiltonian is constant in time.

Proposition 4.2. *Let $z(t)$ be the solution of (4.1) at time t. Further suppose that $\tilde{z}(t)$ is the approximate solution of the reduced system (4.10) in the original coordinate system. Then the error in the Hamiltonian, defined as*

$$\Delta H(t) = |H(z(t)) - H(\tilde{z}(t))|, \tag{4.11}$$

is constant for all $t \in \mathbb{R}$.

Proof. Let φ_t and $\tilde{\varphi}_t$ be the Hamiltonian flow of the original and the reduced system respectively. By definition, $z(t) = \varphi_t(z_0)$ and $y(t) = \tilde{\varphi}_t(y_0)$. Using the definition of the reduced Hamiltonian $\tilde{H}$ and Corollary 2.11 we have

$$H(\tilde{z}(t)) = H(Ay(t)) = \tilde{H}(y(t)) = \tilde{H}(\psi_t(y_0)) = \tilde{H}(y_0) = \tilde{H}(A^+ z_0) = H(AA^+ z_0). \quad (4.12)$$

The error in the Hamiltonian can then be written in terms of z_0 and the symplectic basis A as

$$\Delta H(t) = |H(z_0) - H(AA^+ z_0)| \quad (4.13)$$

□

The following theorems provide a strong indication of the stability of the reduced system.

Definition 4.1. *[15] Consider a dynamical system of the form* $\dot{z} = f(z)$ *and suppose that* z_e *is an equilibrium point for the system so that* $f(z_e) = 0$. z_e *is called nonlinearly stable or Lyapunov stable if, for any* $\epsilon > 0$, *we can find* $\delta > 0$ *such that for any trajectory* ϕ_t, *if* $\|\phi_0 - z_e\|_2 \leq \delta$, *then for all* $0 \leq t < \infty$, *we have* $\|\phi_t - z_e\|_2 < \epsilon$, *where* $\|\cdot\|_2$ *is the Euclidean norm.*

The following proposition, also known as Dirichlet's theorem [15], states a sufficient condition for an equilibrium point to be Lyapunov stable. We refer the reader to [15] for the proof.

Proposition 4.3. *[15] An equilibrium point* z_e *is Lyapunov stable if there exists a scalar function* $W : \mathbb{R}^n \to \mathbb{R}$ *such that* $\nabla W(z_e) = 0$, $\nabla^2 W(z_e)$ *is positive definite, and that for any trajectory* φ_t *defined in the neighborhood of* z_e, *we have* $\frac{d}{dt} W(\varphi_t) \leq 0$. *Here* $\nabla^2 W$ *is the Hessian matrix of* W.

The scalar function W is referred to as the *Lyapunov function*. In the context of the Hamiltonian systems, a suitable candidate for the Lyapunov function is the Hamiltonian function H. The following theorem shows that when H (or $-H$) is a Lyapunov function, then the equilibrium points of the original and the reduced system are Lyapunov stable [1].

Theorem 4.4. *[2] Consider a Hamiltonian system of the form* (4.1) *with a Hamiltonian* $H \in C^2$ *together with the reduced system* (4.10). *Suppose that* z_e *is a strict local minimum of* H. *Furthermore, suppose that* H *(or* $-H$*) is a Lyapunov function satisfying Proposition 4.3. If we can find an open ball neighborhood* S *of* z_e *such that*

Range$(A) \cap S \neq \emptyset$*, then the reduced system* (4.10) *has a stable equilibrium point in* *Range*$(A) \cap S$.

Proof. Assume that H satisfies the conditions in Proposition 4.3 for a Lyapunov function. Since z_e is a local minimum of H, smoothness of H implies that $\nabla_z H(z_e) = 0$, and threfore z_e is a Lyapunov stable point for (4.1). Furthermore, since z_e is a strict local minimum, we can find an open ball S of z_e such that $H(z_e) < H(z)$, for any $z \in \overline{S}$ and $z \neq z_e$, where $\overline{S}$ is the closure of B. Note that $\overline{S}$ is bounded. Since $H \in C^2$ and $\nabla^2 H$ is positive definite at z_e, we also take S to be small enough such that $\nabla^2 H > 0$, for all $z \in S$. Define $c := \inf_{z \in \partial S} H(z)$, where ∂S is the boundary of S. Since H is continues and z_e is a strict minimum we can assume that $H(z) < c$, for all $z \in S$ [1].

Let $S_A = \text{Range}(A) \cap S$. Since $\text{Range}(A)$ is a linear vector space, then S_A is an open set. Furthermore, for any $z \in S_A$, $H(z) < c$.

We now show that $H|_{S_A}$ attains its minimum inside S_A. Let $c_{\min} = \inf_{z \in S_A} H(z)$. c_{min} exists since H has a minimum on S. We can find a sequence $\{H(z_i)\}_{i=1}^{\infty}$, with $z_i \in S_A$, such that $H(z_i) \to c_{\min} < c$. This implies that $z_i \to z_0$, for some $z_0 \in \overline{S_A}$, since $H \in C^2$. Note that $\overline{S_A}$ is bounded since S is bounded. However, z_0 does not belong to ∂S_A since $\inf_{z \in \partial S} W(z) = c > c_{min}$. Therefore $z_0 \in S_A$.

We claim that $y_e = \mathbb{J}_{2k}^T A^T \mathbb{J}_{2n} z_0$ is a stable equilibrium point for the reduced system (4.10). Let $\tilde{H}(y) = H(Ay)$. Note that $\tilde{H}$ attains its local minimum at y_e. Furthermore, $\nabla \tilde{W}(y_e) = 0$. Also we have

$$\nabla^2 \tilde{H} = A^T \nabla^2 H A \tag{4.14}$$

is a positive definite matrix. Finally, since the reduced system is a Hamiltonian system, Corollary 2.11 implies that any trajectory φ_t of (4.10) satisfies $\frac{d}{dt}\tilde{H}(\varphi_t) = 0$. Therefore $\tilde{H}$ is a Lyapunov function for (4.10) and y_e is a stable equilibrium point for (4.10), in the Lyapunov sense.

Similar derivations confirms the theorem when $-H$ is a candidate for a Lyapunov function. □

While the symplectic structure is not guaranteed to be preserved in the reduced systems obtained by the Petrov-Galerkin projection, the reduced system obtained by the symplectic projection guarantees the preservation of the Hamiltonian up to the error (4.11). In the next section we discuss different methods for recovering a symplectic basis.

[1] If there is no such open ball, we can construct a sequence $\{z\}_{i=1}^{\infty}$ such that $z_i \to z_e$ and $H(z_i) = c$, for all i, implying that $H(z_e) = c$ which is a contradiction.

4.2 Proper Symplectic Decomposition

Let S be the snapshot matrix of the Hamiltonian system (4.1). Suppose that a symplectic basis A of size $2n \times 2k$ for the symplectic subspace $\mathcal{A}$ is provided. The *proper symplectic decomposition* (PSD) requires that the error in the symplectic projection onto $\mathcal{A}$ be minimized. Hence, the PSD symplectic basis of size $2k$ is the solution to the minimization problem

$$\begin{aligned} &\underset{A \in \mathbb{R}^{2n \times 2k}}{\text{minimize}} \quad \|S - AA^{+}S\|_F, \\ &\text{subject to} \quad A^T \mathbb{J}_{2n} A = \mathbb{J}_{2k}. \end{aligned} \tag{4.15}$$

Compared to the minimization problem for POD in (3.14), in the above minimization, the orthogonal projection is replaced with a symplectic projection AA^{+}. At first, the minimization looks similar to the one obtained by POD. However, it is well known that symplectic bases are not generally orthogonal, and therefore not norm bounded. This means that numerical errors may become dominant in a symplectic projection [61] which makes the minimization (4.15) a harder problem than (3.14).

As the optimization problem (4.15) is nonlinear, the direct solution is usually expensive. A simplified version of this optimization can be found in [81], but there is no guarantee that the method provides a near optimal basis.

Finding eigen-spaces of Hamiltonian and symplectic matrices is studied in the context of optimal control problems [11, 13, 109, 20] and model reduction of Riccati equations [13], where also an SVD-like decomposition for Hamiltonian and symplectic matrices has been proposed [111]. The computation of Lagrangian subspaces of large scale Hamiltonian matrices using a CS-decomposition is presented in [71, 70]. However, the computation of a large snapshot matrix and the use of the methods to compute its eigen-spaces, is usually computationally demanding. Also, these methods generally do not guarantee the construction of a well-conditioned symplectic basis.

In Section 4.2.1 we briefly outline non-direct methods for finding solutions to (4.15), proposed in [81], assuming a specific structure for A. In Section 4.3 we propose a greedy approach for the symplectic basis generation.

4.2.1 SVD Based Methods for Symplectic Basis Generation

The methods presented in this section are taken from [81].

Cotangent lift: Suppose that A is of the form

$$A = \begin{pmatrix} \Phi & 0 \\ 0 & \Phi \end{pmatrix}, \tag{4.16}$$

where $\Phi \in \mathbb{R}^{n\times k}$ is an orthonormal matrix. It is easy to check that A is a symplectic matrix, i.e., $A^T \mathbb{J}_{2n} A = \mathbb{J}_{2k}$. The construction of A suggests that the range of Φ should cover both the potential and the momentum spaces of the Hamiltonian problem. Hence, we can construct A by forming the combined snapshot matrix

$$S_{\text{comb}} = [q_1, \ldots, q_n, p_1, \ldots, p_n], \qquad z_i = (q_i^T, p_i^T)^T, \tag{4.17}$$

and define $\Phi = [u_1, \ldots, u_k]$, where u_i is the i-th left singular vector of S_{comb}. It is shown in [81] that among all symplectic bases of the form (4.16), the cotangent lift minimizes the projection error in L^2.

Complex SVD: Suppose instead that A takes the form

$$A = \begin{pmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{pmatrix}, \tag{4.18}$$

where $\Phi, \Psi \in \mathbb{R}^{n\times k}$ are real matrices of size $n \times k$ satisfying

$$\Phi^T\Phi + \Psi^T\Psi = I_k, \quad \Phi^T\Psi = \Psi^T\Phi. \tag{4.19}$$

It can be checked that A forms a symplectic matrix [69]. To construct A we first define the complex snapshot matrix

$$S_{\text{comp}} = [q_1 + ip_1, \ldots, q_n + ip_n]. \tag{4.20}$$

Each left singular vector of S_{comp} now takes the form $u_m = r_j + is_j$. We define

$$\Phi = [r_1, \ldots, r_k], \quad \Psi = [s_1, \ldots, s_k]. \tag{4.21}$$

One can easily check that (4.19) is satisfied since the matrix of singular vectors is unitary. It is shown in [81] that among all symplectic bases of the form (4.18) the complex SVD minimizes the projection error in L^2.

4.3 The Greedy Approach to Symplectic Basis Generation

We discussed the greedy generation of a reduced basis in Section 3.3. In this section we generalize the method to construct a symplectic basis.This method adds the two best possible basis vectors to the symplectic basis to enhance overall accuracy measured in L^∞. In contrast to the cotangent lift and the complex SVD methods, the greedy

approach does not require the symplectic basis to have a specific structure. This typically results in a more compact basis and/or a more accurate reduced systems. For parametric problems, the greedy approach only requires one numerical solution to be computed per iteration, hence saving substantial computational cost in the offline stage.

The orthonormalization step is an essential step in most greedy approaches for basis generation in the context of model reduction [52, 86]. However common orthonormalization methods, e.g. the QR method, destroy the symplectic structure of the original system [20]. Here we use a variant of the QR method known as the SR [95] method that is based on the symplectic GS method, introduced in Corollary 2.8.

As discussed in Section 2.4, any finite dimensional symplectic linear vector space can be equipped with a canonical basis. Furthermore, Corollary 2.8 and Proposition 2.12 provides an iterative process for constructing an ortho-symplectic basis [95]. To briefly describe the SR method, suppose $E_{2k} = \{e_i, T^{-1}(e_i)\}_{i=1}^{k}$ is a given ortho-symplectic basis with respect to the Euclidean inner product, where T is the transformation defined in Proposition 2.12. For the Euclidean inner product, it is easily verified that $T^{-1}(e_{k+1}) = \mathbb{J}_{2n}^T e_{k+1}$. Furthermore, let A_{2k} be the matrix that contains these vectors in its columns and let z be a given vector such that $z \notin \mathcal{A}_{2k}$, the span space of E_{2k}. We first remove the contribution of E_{2k} from z to obtain

$$\tilde{z} = z - P^{\text{symp}}_{\mathcal{A}_{2k}}(z). \tag{4.22}$$

If we introduce $e_{k+1} = \tilde{z}/\|\tilde{z}\|_2$, it is easily checked that e_{k+1} is also orthogonal to A_{2k} with respect to the Euclidean inner product. Therefore, span$\{e_1, \ldots, e_{k+1}\}$ forms a Lagrangian subspace. Furthermore, the basis $E_{2k+2} = E_{2k} \cup \{e_{k+1}, \mathbb{J}_{2n}^T e_{k+1}\}$ forms an ortho-symplectic basis. Finally we can assemble the matrix for the reduced basis of size $2k+2$ as

$$A_{2k+2} = [e_1, \ldots, e_{k+1}, \mathbb{J}_{2n}^T e_1, \ldots, \mathbb{J}_{2n}^T e_{k+1}]. \tag{4.23}$$

Note that the SR method is chosen due to its simplicity. It can be replaced with backward stable routines such as the isotropic Arnoldi or the isotropic Lanczos methods [72], if required.

A key element of the greedy algorithm is the availability of an error indicator which efficiently evaluates the error associated with the reduced model [52]. In the framework of symplectic model reduction for a parametric Hamiltonian system, one possible candidate is the error in the Hamiltonian (4.11). Correctly approximating symplectic systems relies on preservation of the Hamiltonian, hence the error in the Hamiltonian arises as a natural choice. Moreover, since the error in the Hamiltonian depends on the initial condition and the reduced symplectic basis, evaluation of the error does

not require the time integration of the full system. On the other hand, the error in the Hamiltonian fails to identify the best snapshot when $z_0 \in E_{2k}$, where z_0 is the initial condition in (4.1).

Suppose that a $2k$-dimensional ortho-symplectic basis E_{2k} is generated at the k-th step of the greedy method and we seek to enrich it by two additional vectors. Using the error in the Hamiltonian (4.13) we search the parameter space to identify the value that maximizes the error in the Hamiltonian

$$\mu_{k+1} := \operatorname*{argmax}_{\mu \in \mathbb{P}^{\Delta}} \Delta H(\mu). \tag{4.24}$$

Here, $\mathbb{P}^{\Delta} \subset \mathbb{R}^d$ is the discretized parameter space. The goal is to approximate the Hamiltonian function as well as possible. We compute the temporal snapshots of (4.1) with respect to μ_{k+1} and form the temporal snapshot matrix

$$S_{t,\mu_{k+1}} = [z(t_i; \mu_{k+1})]_{i=1}^{N_t}. \tag{4.25}$$

The next basis vector is the snapshot that maximises the projection error

$$z := \operatorname*{arg\,max}_{s \in S_{t,\mu_{k+1}}} \|s - P^{\text{symp}}_{\mathcal{A}_{2k}}(s)\|_2. \tag{4.26}$$

Finally, we update the basis as

$$e_{k+1} = \tilde{z}, \quad E_{2k+2} = E_{2k} \cup \{e_{k+1}, \mathbb{J}_{2n}^T e_{k+1}\}, \tag{4.27}$$

where $\tilde{z}$ is the vector obtained after symplectic orthonormalization of z with respect E_{2k}. Finally A_{2k+2} can be assembled according to (4.23).

Note that the greedy-POD method, introduced in Section 3.3, cannot be directly applied in the symplectic setting since the union of two disjoint symplectic bases (as apposed to orthogonal bases) is not guaranteed to be symplectic.

Since the maximization over the entire parameter space $\mathbb{P}$ is impossible, we discretize the parameter set into a grid with N points: $\mathbb{P}_N = \{\mu_1, \ldots, \mu_N\}$. However, since the selection of parameters only require the evaluation of the error in the Hamiltonian and not time integration of the original system, then $\mathbb{P}_N$ can be chosen to be rich.

We summarize the greedy algorithm for the generation of a symplectic basis in Algorithm 4.1.

Algorithm 4.1 the symplectic greedy for extending a symplectic reduced basis

Input: parameter space $\mathbb{P}^{\Delta}$, error indicator function η , symplectic reduced basis A_{2k}.

1: find $\mu^* := \arg\max_{\mu\in\mathbb{P}^{\Delta}} \eta(\mu)$.
2: compute the temporal snapshots S_{t,μ^*}.
3: Find the snapshot with maximum projection error

$$z := \arg\max_{s\in S_{t,\mu_k}} \|s - A_{2k}{A_{2k}}^{+}s\|_2.$$

4: Apply symplectic orthonormalization on z to obtain e_{k+1}.
5: Assemble $A_{2k+2} = [e_1, \ldots, e_{k+1}, \mathbb{J}_{2n}^T e_1, \ldots, \mathbb{J}_{2n}^T e_{k+1}]$.

Output: symplectic reduced basis A_{2k+2}.

4.4 Convergence of the Greedy Method

In Section 3.3 we discussed that the conventional greedy basis generation, in the strong sense, is exponentially accurate. In this section we show that the symplectic greedy process, with the symplectic projection error as an error estimator, maintains this property.

Suppose that we are given a compact subset S of $\mathbb{R}^{2n}$. Our intention is to find a set of vectors $E_{2k} = \{e_i, f_i\}_{i=1}^k$ such that E_{2k} forms an orthosymplectic basis and any $s \in S$ is well approximated by elements of the subspace $\mathcal{A}_{2k} = \text{span}(E_{2k})$. The greedy process using the projection error for generating basis vectors e_i and f_i is as follows. In the initial step we pick e_1 such that $\|e_1\|_2 = \max_{s\in S} \|s\|_2$. And define $f_1 = \mathbb{J}_{2n}^T e_1$. As discussed in Section 4.3, $E_2 = \{e_1, f_1\}$ is orthosymplectic, so E_2 is the first subspace that approximates elements of S. In the k-th step of the greedy method, suppose we have a basis $E_{2k} = \{e_1, \ldots, e_k, f_1, \ldots, f_k\}$, with A_{2k} the matrix that contain these vectors in its column. Let $P^{\text{symp}}_{\mathcal{A}_{2k}}$ be the symplectic projection onto $\mathcal{A}_{2k}$ and define

$$\sigma_{2k}(s) := \|s - P^{\text{symp}}_{\mathcal{A}_{2k}}(s)\|_2, \tag{4.28}$$

as the projection error. Moreover we denote by σ_{2k} the maximum approximation error of S using elements in $\text{span}(A_{2k})$ as

$$\sigma_{2k} := \max_{s\in S} \sigma_{2k}(s). \tag{4.29}$$

The next set of basis vectors in the greedy selection are

$$e_{k+1} := \arg\max_{s\in S} \sigma_{2k}(s), \quad f_{k+1} := \mathbb{J}_{2n}^T e_{k+1}. \tag{4.30}$$

We emphasize that the sequence of basis vectors generated by the greedy is in general

not unique [86, 52].

To estimate the quality of the reduced subspace, it is natural to compare it with the best possible $2k$-dimensional subspace in the sense of the minimum projection error (not necessarily in the symplectic sense). For this we use the Kolmogorov n-width [63, 83], see Definition 3.3. To recall, for a given subspace $\mathcal{W}_n$,

$$\operatorname{dist}(S, \mathcal{W}_n) = \sup_{s \in S} \operatorname{dist}(s, \mathcal{W}_n), \tag{4.31}$$

measures the worst possible projection error of elements in S onto $\mathcal{W}_n$. Hence the Kolmogorov n-width quantifies how well S can possibly be approximated by an n-dimensional subspace.

We seek to show that the decay of σ_{2k}, obtained by the greedy algorithm, has the same rate as $d_{2k}(S)$, i.e., the greedy method provides the best possible accuracy attained by a $2k$-dimensional subspace.

We start by $\mathbb{J}_{2n}$-orthogonalizing the vectors provided by the greedy algorithm as

$$\begin{aligned} &\xi_1 = e_1, && \bar{\xi}_1 = \mathbb{J}_{2n}^T \xi_1, \\ &\xi_i = e_i - P_{2(i-1)}(e_i), && \bar{\xi}_i = \mathbb{J}_{2n}^T, \xi_i \quad i = 2, 3, \dots \end{aligned} \tag{4.32}$$

The projection of a vector $s \in S$ onto $\mathcal{A}_{2k}$ can be written using the symplectic basis as

$$P^{\text{symp}}_{\mathcal{A}_{2k}}(s) = \sum_{i=1}^{k} \left(\alpha_i(s) \xi_i + \bar{\alpha}_i(s) \bar{\xi}_i \right), \tag{4.33}$$

where $\alpha_i(s)$ and $\bar{\alpha}_i(s)$ for $i = 1, \dots, k$ are the expansion coefficients

$$\alpha_i(s) = -\frac{\Omega(\bar{\xi}_i, s)}{\Omega(\xi_i, \bar{\xi}_i)}, \quad \bar{\alpha}_i(s) = \frac{\Omega(\xi_i, s)}{\Omega(\xi_i, \bar{\xi}_i)}, \tag{4.34}$$

for any $s \in S$. Since $\bar{\xi}_i$ is $\mathbb{J}_{2n}$-orthogonal to $\mathcal{A}_{2k-2}$ we have

$$\begin{aligned} |\alpha_i(s)| &= \frac{|\Omega(\bar{\xi}_i, s)|}{|\Omega(\xi_i, \bar{\xi}_i)|} = \frac{|\Omega(\bar{\xi}_i, s - P^{\text{symp}}_{\mathcal{A}_{2k-2}}(s))|}{|\Omega(\xi_i, \bar{\xi}_i)|} \leq \frac{\|\bar{\xi}_i\|_2 \|s - P^{\text{symp}}_{\mathcal{A}_{2k-2}}(s)\|_2}{\|\xi_i\|_2 \|\bar{\xi}_i\|_2} \\ &= \frac{\|s - P^{\text{symp}}_{\mathcal{A}_{2k-2}}(s)\|_2}{\|e_i - P^{\text{symp}}_{\mathcal{A}_{2k-2}}(e_i)\|_2} \leq 1. \end{aligned} \tag{4.35}$$

Here, we use the fact that $|\Omega(\xi_i, \bar{\xi}_i)| = \|\xi_i\|_2^2 = \|\bar{\xi}_i\|_2^2$ with the last inequality following from the greedy algorithm which maximizes e_i. Similarly we deduce that $|\bar{\alpha}_i(s)| \leq 1$.

We write

$$\xi_j = \sum_{i=1}^{j} \left(\mu_i^j e_i + \gamma_i^j f_i\right), \quad \bar{\xi}_j = \sum_{i=1}^{j} \left(\lambda_i^j e_i + \eta_i^j f_i,\right), \quad j = 1, 2, \ldots \tag{4.36}$$

with

$$\begin{aligned} &\mu_j^j = 1, \quad \gamma_j^j = 0, \\ &\mu_i^j = \sum_{l=i}^{j-1} \left(-\alpha_l(f_j)\mu_i^l + \bar{\alpha}_l(f_j)\gamma_i^l\right), \quad \gamma_i^j = \sum_{l=i}^{j-1} \left(-\alpha_l(f_j)\gamma_i^l + \bar{\alpha}_l(f_j)\mu_i^l\right), \\ &\lambda_i^j = -\gamma_i^j, \quad \eta_i^j = \mu_i^j, \end{aligned} \tag{4.37}$$

for $j = 2, 3, \ldots$. By induction and using the bound in (4.35) we recover

$$\mu_i^j, \gamma_i^j, \lambda_i^j, \eta_i^j \leq 3^{j-i}, \quad \text{for } j \geq i. \tag{4.38}$$

Now let $2k$ be the dimension of the desired reduced space. By the definition of Kolmogorov n-width we observe that for any $\theta > 1$ we can find a subspace $\mathcal{W}_{2k}$ such that $\text{dist}(S, \mathcal{W}_{2k}) \leq \theta d_{2k}(S, \mathbb{R}^n)$. Hence we can find vectors $v_1, \ldots, v_k, u_1, \ldots, u_k \in \mathcal{W}_{2k}$ such that

$$\begin{aligned} &\|e_i - v_i\|_2 \leq \theta d_{2k}(S, \mathbb{R}^n), \\ &\|f_i - u_i\|_2 \leq \theta d_{2k}(S, \mathbb{R}^n). \end{aligned} \tag{4.39}$$

Now we construct a set of $2(k+1)$ new vectors

$$\zeta_j = \sum_{i=1}^{k+1} \mu_i^j v_i + \gamma_i^j u_i, \quad \bar{\zeta}_j = \sum_{i=1}^{k+1} \lambda_i^j v_i + \eta_i^j u_i. \tag{4.40}$$

for $j = 1, \ldots, k+1$. Note that since u_i and v_i belong to $\mathcal{W}_{2k}$ so does their linear combination including all ζ_j and $\bar{\zeta}_j$. We can use the inequality (4.38) to write

$$\|\xi_i - \zeta_i\|_2 \leq 3^i \theta d_{2k}(S, \mathbb{R}^n), \quad \|\bar{\xi}_i - \bar{\zeta}_i\|_2 \leq 3^i \theta d_{2k}(S, \mathbb{R}^n). \tag{4.41}$$

Moreover since $\mathcal{W}_{2k}$ is of dimension $2k$ we find κ_i, $i = 1, \ldots, 2(k+1)$ such that

$$\sum_{i=1}^{2(k+1)} \kappa_i^2 = 1, \quad \sum_{i=1}^{k+1} \kappa_i \zeta_i + \sum_{i=1}^{k+1} \kappa_{i+k+1} \bar{\zeta}_i = 0. \tag{4.42}$$

We have

$$\begin{aligned} \left\| \sum_{i=1}^{k+1} \kappa_i \xi_i + \sum_{i=1}^{k+1} \kappa_{i+k+1} \bar{\xi}_i \right\|_2 &= \left\| \sum_{i=1}^{k+1} \kappa_i (\xi_i - \zeta_i) + \sum_{i=1}^{k+1} \kappa_{i+k+1} (\bar{\xi}_i - \bar{\zeta}_i) \right\|_2 \\ &\leq 2 \cdot 3^{k+1} \sqrt{2(k+1)} \theta d_{2k}(S, \mathbb{R}^n). \end{aligned} \tag{4.43}$$

We know there exists $1 \leq j \leq 2k+2$ such that $\kappa_j > 1/\sqrt{2(k+1)}$. Without loss of generality let us assume that $j \leq k+1$. This yields

$$\left\| \xi_j + \kappa_j^{-1} \sum_{i=1, i\neq j}^{k+1} \kappa_i \xi_i + \kappa_j^{-1} \sum_{i=1}^{k+1} \kappa_{i+k+1} \bar{\xi}_i \right\|_2 \leq 4 \cdot 3^{k+1}(k+1)\theta d_{2k}(S, \mathbb{R}^n). \tag{4.44}$$

Define $c = \kappa_j^{-1} \sum_{i=1, i\neq j}^{k+1} \kappa_i \xi_i + \kappa_j^{-1} \sum_{i=1}^{k+1} \kappa_{i+k+1} \bar{\xi}_i$. Using that $\mathbb{J}_{2n}^T c$ is $\mathbb{J}_{2n}$-orthogonal to ξ_j we recover

$$\begin{aligned} \|\xi_j\|_2 \leq \|\xi_j\|_2 + \|c\|_2 &= \Omega(\xi_j, \mathbb{J}_{2n}^T \xi_j) + \Omega(c, \mathbb{J}_{2n}^T c) \\ &= \Omega(\xi_j, \mathbb{J}_{2n}^T \xi_j) + \Omega(c, \mathbb{J}_{2n}^T c) + \Omega(\xi_j, \mathbb{J}_{2n}^T c) + \Omega(c, \mathbb{J}_{2n}^T \xi_j) \\ &= \Omega(\xi_j + c, \mathbb{J}_{2n}^T (\xi_j + c)) = \|\xi_j + c\|_2 \end{aligned} \tag{4.45}$$

Combining this with (4.44) yields

$$\|\xi_j\|_2 \leq 4 \cdot 3^{k+1}(k+1)\theta d_{2k}(S, \mathbb{R}^n). \tag{4.46}$$

Finally using the definition of ξ_j for all $s \in S$ we have

$$\|s - P_{2(j-1)}(s)\|_2 \leq \|f_j - P_{2(j-1)}(f_j)\|_2 = \|\xi_j\|_2 \leq 4 \cdot 3^{k+1}(k+1)\theta d_{2k}(S, \mathbb{R}^n) \tag{4.47}$$

Hence, for any given $\lambda > 1$

$$\|s - P_{2k}(s)\|_2 \leq \|s - P_{2(j-1)}(s)\|_2 \leq 4 \cdot 3^{k+1}(k+1)\theta d_{2k}(S, \mathbb{R}^n). \tag{4.48}$$

This establishes the following theorem.

Theorem 4.5. *Let S be a compact subset of $\mathbb{R}^{2n}$ with exponentially small Kolmogorov n-width $d_k \leq c\exp(-\alpha k)$ with $\alpha > \log 3$. Then there exists $\beta > 0$ such that the symplectic subspaces A_{2k} generated by the greedy algorithm provide exponential approximation properties such that*

$$\|s - P_{\mathcal{A}_{2k}}^{symp}(s)\|_2 \leq C\exp(-\beta k) \tag{4.49}$$

for all $s \in S$ and some $C > 0$.

Note the convergence property in Theorem 4.5 does not in general hold when the error in the Hamiltonian, introduced in (4.11), is used as an error indicator. However, the evaluation of the projection error is often expensive. Therefore, the error in the Hamiltonian can be a cheap surrogate. A numerical comparison between the two error estimators is presented in Section 4.6.3 and also in [89].

4.5 Symplectic Discrete Empirical Interpolation Method (SDEIM)

In Section 3.5 we discussed the computational challenges associated with evaluating nonlinear terms in the context of RB methods. In general, conventional methods, e.g. EIM, destroy the symplectic structure of a Hamiltonian system. In this section we discuss how such methods can be modified to accelerate the evaluation of nonlinear terms, while preserving the symplectic structure.

Consider the Hamiltonian system (4.1) and its reduced system (4.10) equipped with a symplectic transformation A. One can split the Hamiltonian function $H = H_1 + H_2$ such that $\nabla H_1 = Lz$ and $\nabla H_2 = g(z)$, where L is a constant matrix in $\mathbb{R}^{2n\times 2n}$ and g is a nonlinear function. Substituting this in (4.10) yields

$$\frac{d}{dt}y = \mathbb{J}_{2k}\underbrace{A^T L A}_{\tilde{L}}y + A^+\mathbb{J}_{2n}g(Ay). \tag{4.50}$$

Here, we used the fact that $\nabla_y H_1 = A^T L A$ and Proposition 4.1. As discussed in Section 3.5, the complexity of evaluating the nonlinear term still depends on n, the size of the original system. To overcome this computational bottleneck we consider the DEIM approximation for evaluating the nonlinear function g as

$$\frac{d}{dt}y = \mathbb{J}_{2k}\tilde{L}y + \underbrace{A^+\mathbb{J}_{2n}V(P^TV)^{-1}P^Tg(Ay)}_{N(y)}, \tag{4.51}$$

where V is a basis for the nonlinear snapshots $S_g = [g(z_i)]_{i=1}^{N_t}$ and P is the index matrix for interpolation in Algorithm 3.5. For a general choice of V the system (4.51) is not guaranteed to be a Hamiltonian system, impacting long time accuracy and stability. However, we can guarantee that (4.51) is a Hamiltonian system by choosing $V = (A^+)^T$. To see this, note that the system (4.51) is a Hamiltonian system if and only if $N(y) = \mathbb{J}_{2k}\nabla_y\bar{H}(y)$ for some scalar function $\bar{H}$. Also we have

$$g(Ay) = \nabla_z H_2(z) = (A^+)^T\nabla_y H_2(Ay), \tag{4.52}$$

Substituting this into N we obtain

$$N(y) = A^+\mathbb{J}_{2n}V(P^TV)^{-1}P^T(A^+)^T\nabla_y H_2(Ay). \tag{4.53}$$

Taking $V = (A^+)^T$ yields

$$N(y) = A^+\mathbb{J}_{2n}(A^+)^T\nabla_y H_2(Ay) = \mathbb{J}_{2k}\nabla_y H_2(Ay), \tag{4.54}$$

since $(A^+)^T$ is a symplectic matrix. Hence, $V = (A^+)^T$ is a sufficient condition for (4.51) to be Hamiltonian.

Regarding the construction of the projection space, suppose that we have already constructed a symplectic basis $A = \{e_1, \dots, e_k, f_1, \dots f_k\}$ using the greedy algorithm. Note that $(A^+)^T$ is a symplectic basis and $(A^+)^+ = A$. Thus, we can move between these two symplectic bases by simply using the transpose operator and the symplectic inverse operator. Let S_g be the nonlinear snapshots. We form $(A^+)^T = \{e'_1, \dots, e'_k, f'_1, \dots, f'_k\}$ and use a greedy approach to add new basis vectors to $(A^+)^T$. At the i-th iteration of the symplectic DEIM, we use $(A^+)^T$ to approximate elements in S_g and choose the vector that maximizes the error as the next basis vector

$$s^* := \underset{s \in S_g}{\operatorname{argmax}} \|s - (A^+)^T A^+ s\|_2. \tag{4.55}$$

After applying the symplectic Gram-Schmidt on s^*, we update $(A^+)^T$ using vectors

$$e'_{k+i+1} = \frac{s^*}{\|s^*\|_2}, \quad f'_{k+i+1} = \mathbb{J}_{2n}^T e'_{k+i+1}. \tag{4.56}$$

Finally when $(A^+)^T$ approximates elements S_g with the desired accuracy, we transpose and symplectically invert $(A^+)^T$ to obtain A. We summarize the symplectic DEIM algorithm in Algorithm 4.2.

Algorithm 4.2 the symplectic DEIM for extending a symplectic basis A_{2k}

Input: nonlinear snapshots S_g, symplectic reduced basis A_{2k}.

1: compute $B_{2k} := (A_{2k}^+)^T = [e'_1, \dots, e'_k, \mathbb{J}_{2n}^T e'_1, \dots, \mathbb{J}_{2n}^T e'_k]$
2: find $s^* := \underset{s \in S_g}{\operatorname{argmax}} \|s - BB^+ s\|_2$.
3: Apply symplectic orthonormalization on s^* to obtain e'_{k+1}.
4: Assemble $B_{2k+2} = [e'_1, \dots, e'_{k+1}, \mathbb{J}_{2n}^T e'_1, \dots, \mathbb{J}_{2n}^T e'_{k+1}]$.
5: Compute $A_{2k+2} = (B^+)^T$.

Output: symplectic reduced basis A_{2k+2}.

When the symplectic reduced basis is an ortho-symplectic basis of the form $A = [e_1, \dots, e_k, \mathbb{J}_{2k}^T e_1, \dots, \mathbb{J}_{2k}^T e_k]$, Proposition 4.1 suggests that $(A^+)^T = A$. Therefore, A can be directly enriched with nonlinear snapshots. This is also suggested in [81]. However, not all ortho-symplectic bases are of this form. Given a proper vector selection procedure, the symplectic DEIM can be modified to construct an accurate symplectic basis for any symplectic basis. This feature is exploited in Section 5.5, where the reduced symplectic basis is no longer orthonormal with respect to the Euclidean inner-product.

Note that Proposition 4.1 implies that $(A^+)^T = A$, therefore it suffices to ensure that A is also a basis for S_g, the nonlinear snapshots. Suppose that we have already constructed a symplectic basis $E = \{e_1, \dots, e_k, f_1, \dots f_k\}$ using the symplectic greedy in Algorithm 4.1. We seek to enrich it with the nonlinear snapshots. We can use the

projection error to identify the nonlinear snapshot that is worst approximated by $\mathcal{A} = \text{span}(E)$ as

$$s^* := \underset{s \in S_g}{\arg\max} \|s - AA^+ s\|_2. \tag{4.57}$$

We then symplectically ortho-normalize s^* to obtain e_{k+1} and $f_{k+1} = \mathbb{J}_{2n}^T e_{k+1}$ and add them to E. We can continue this process until E approximates the nonlinear snapshots with the desired accuracy. Therefore, we may simply call Algorithm 4.1 while passing a symplectic basis A with the nonlinear snapshots S_g as its arguments.

When using an implicit time integration scheme we face inefficiencies when evaluating the Jacobian of nonlinear terms, as discussed in Section 3.5. We recall that a key ingredient to a fast approximation of the Jacobian is that the interpolating index matrix P, obtained in the DEIM approximation, commutes with the nonlinear function. Nonlinear terms in Hamiltonian systems often take the from

$$g(z) = g(q,p) = \begin{pmatrix} g_1(q_1,p_1) \\ g_2(q_2,p_2) \\ \vdots \\ g_{2n}(q_n,p_n) \end{pmatrix}, \quad g_i : \mathbb{R}^2 \to \mathbb{R}, \quad 1 \leq i \leq n. \tag{4.58}$$

Thus, the interpolating index matrix, obtained by Algorithm 3.5 does not necessarily commute with the function g. To overcome this, when index $\mathfrak{p}_i$ with $\mathfrak{p}_i \leq n$ or $\mathfrak{p}_i > n$ is chosen in Algorithm 3.5 we also include $\mathfrak{p}_i + n$ or $\mathfrak{p}_i - n$, respectively. Simple calculations verify that g and P then commute.

In case g is not of the form (4.58) one can use MDEIM [23, 79] to accelerate the assembly of the Jacobian matrix.

4.6 Numerical Results

In this section, we illustrate the performance of the greedy generation of a symplectic basis. The parametric linear wave equation is considered to compare SVD based methods with the greedy method. The symplectic model reduction of nonlinear systems is then illustrated by considering the parametric nonlinear Schrödinger equation. Finally we discuss the numerical convergence of the greedy method introduced in Algorithm 4.1.

4.6.1 Parametric Linear Wave Equation

Consider the parametric linear wave equation

$$\begin{cases} u_{tt}(x,t,\omega) = \kappa(\omega) u_{xx}(x,t,\omega), \\ u(x,0) = u^0(x), \end{cases} \tag{4.59}$$

where x belongs to a one-dimensional torus of length L, $\omega = (\omega_1, \ldots, \omega_4)$ and

$$\kappa(\omega) = c^2 \left(\sum_{l=1}^{4} \frac{1}{l^2} \omega_l \right). \tag{4.60}$$

Here $\omega_l \in [0,1]$ for $l = 1, \ldots, 4$ and $c \in \mathbb{R}$ is a constant number. By rewriting (4.59) in canonical form, using the change of variable $q = u$ and $\partial q / \partial t = p$, we obtain the symplectic form

$$\begin{cases} q_t(x,t,\omega) = p(x,t,\omega), \\ p_t(x,t,\omega) = \kappa(\omega) q_{xx}(x,t,\omega), \end{cases} \tag{4.61}$$

with the associated Hamiltonian

$$H(q,p,\omega) = \frac{1}{2} \int_0^L p^2 + \kappa(\omega) q_x^2 \, dx. \tag{4.62}$$

We discretize the torus into N equidistant points and define $\Delta x = L/N$, $x_i = i\Delta x$, $q_i = q(t, x_i, \omega)$ and $p_i = p(t, x_i, \omega)$ for $i = 1, \ldots, N$. Furthermore, we discretize (4.61) using a standard central finite differences scheme to obtain

$$\frac{d}{dt} z = \mathbb{J}_{2N} L z, \tag{4.63}$$

where $z = (q, \ldots, q_N, p_q, \ldots, p_n)^T$ and

$$L = \begin{pmatrix} I_n & 0_N \\ 0_N & \kappa(\omega) D_{xx} \end{pmatrix}, \tag{4.64}$$

with D_{xx} the central finite differences matrix operator. The discrete Hamiltonian can finally be written as

$$H_{\Delta x}(z) = \frac{\Delta x}{2} \sum_{i=1}^{N} \left(p_i^2 + \kappa(\omega) \frac{(q_{i+1} - q_i)^2}{2\Delta x^2} + \kappa(\omega) \frac{(q_i - q_{i-1})^2}{2\Delta x^2} \right). \tag{4.65}$$

The initial condition is given by

$$q_i(0) = h(10 \times |x_i - \frac{1}{2}|), \quad p_i = 0, \quad i = 1, \ldots, N \tag{4.66}$$

where $h(s)$ is the cubic spline function

$$h(s) = \begin{cases} 1 - \frac{3}{2}s^2 + \frac{3}{4}s^3, & 0 \leq s \leq 1, \\ \frac{1}{4}(2-s)^3, & 1 < s \leq 2, \\ 0, & s > 2. \end{cases} \tag{4.67}$$

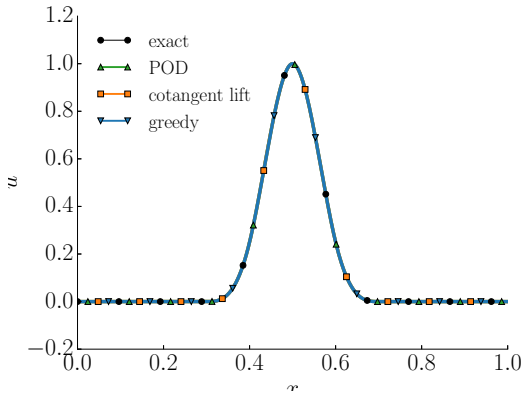
This will result in waves propagating in both directions on the torus.

For numerical time integration we use the Störmer-Verlet scheme (2.36). As the Hamiltonian is *separable*, i.e. $H(q,p) = U(q) + K(p)$, the Störmer-Verlet scheme becomes an explicit time stepping scheme. The high-fidelity system uses the following parameter set

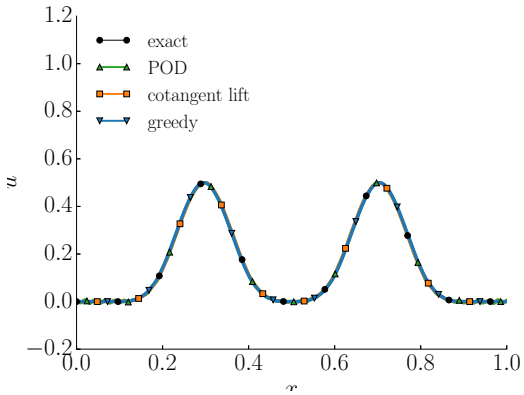
Domain length	$L = 1$
No. grid points	$N = 500$
Space discretization size	$\Delta x = 0.002$
Time discretization size	$\Delta t = 0.01$
Wave speed	$c^2 = 0.1$

We compare the reduced system obtained by the greedy algorithm with the methods based on SVD. To generate snapshots, we discretize the parameter space $[0,1]^4$ into in total of 5^4 equidistant grid points. For the SVD based methods and POD, snapshots are gathered in the snapshot matrices S, S_{comb} and S_{comp}, respectively, and the SVD is performed to construct the reduced basis. The greedy method is applied following Algorithm 4.1; as input, the tolerance for the error in the Hamiltonian is set to $\delta = 5 \times 10^{-3}$. All reduced systems are taken to have an identical size ($k = 80$ for POD and $k = 40$ for the symplectic methods). We use the Störmer-Verlet scheme for symplectic methods and a second order Runge-Kutta method for the POD. The choice of different time integration routines is justified by the fact that the POD destroys the canonical form of the original equations and a symplectic integrator cannot be applied. One can alternatively use separate reduced subspaces for the potential and the momentum spaces, which, however, is not a standard model reduction approach and requires further analysis. Finally we transform the reduced solution into the high-fidelity space for illustration purposes.

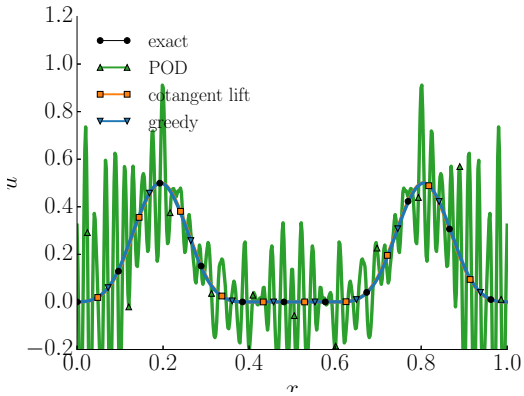
The cost is reduced by 50% in the offline stage when using the greedy method as compared to SVD-based methods (cotangent lift and complex SVD method). This is because the SVD-based methods require time integration of the full system for all discrete parameter points, while the greedy method picks a number of parameters from the parameter space.



(a) $t = 0$

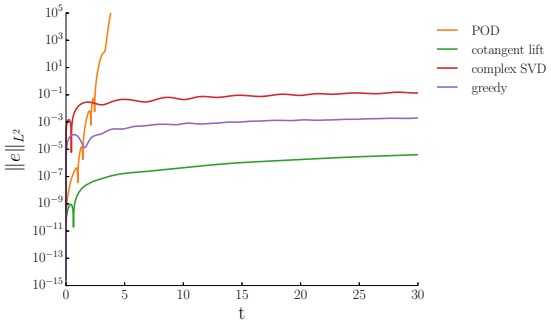


(b) $t = 1$

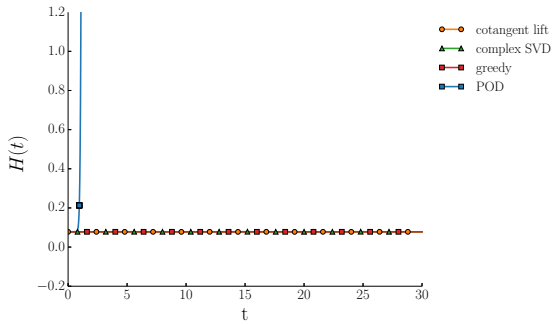


(c) $t = 2$

Figure 4.1 – The solution q at $t = 0$, $t = 1$ and $t = 2$ of the linear wave equation for parameter value $c = 0.1019$ different from training parameters. Here, the solution of the full system together with the solution of the POD, cotangent lift, complex SVD and the greedy reduced system is shown.



(a)



(b)

Figure 4.2 – (a) The L^2-error between the solution of the full system and the reduced system for different model reduction methods for $t \in [0, 30]$. (b) Plot of the Hamiltonian function for $t \in [0, 30]$.

Figure 4.1.(a) shows the solution of the linear wave equation for parameter values $(\omega_1, \omega_2, \omega_3, \omega_4) = (0.8456, 0.1320, 0.9328, 0.5809)$ or $\kappa(\omega) = 0.1019$, chosen to be different from training parameters, at $t = 0$, $t = 1$ and $t = 2$. While we see instability and divergence from the exact solution for the POD reduced system, the symplectic methods provide a good approximation of the full model.

The decay of the singular values for the POD are shown in Figure 4.5.(a). The decay of the singular values suggests that a low dimensional solution manifold indeed exists. However, since the linear subspace, constructed by the POD, is not symplectic, we observe blow up of the Hamiltonian function in Figure 4.2.(b) and the instability of the solution in Figure 4.1. The symplectic methods (using a reduced basis of the same size as POD) preserve the Hamiltonian function as shown in Figure 4.2.(b).

Figure 4.2.(c) shows the L^2-error between the solution of the full model and the reduced systems constructed by different methods. We note that the error for the POD reduced system rapidly increases, confirming the instability of the reduced system. Furthermore, the symplectic methods provide a better approximation since the geometric structure of the original system is preserved. Although the greedy method constructs a basis almost twice as fast compared to the SVD-based methods and the redulting basis is not guaranteed to be optimal in L^2, its accuracy is comparable. The cotangent lift method provides a more accurate solution, on the other hand the cotangent lift basis (4.16) takes a less general form and is usually computationally more demanding than the greedy method.

For complex systems, were the solution of the full system is expensive, and for high dimensional parameter domains, POD-based methods become impractical [52, 86]. However, the greedy method requires substantially fewer (proportional to the size of the reduced basis) evaluation of the time integration of the original system.

4.6.2 Nonlinear Schrödinger Equation

Let us consider the one-dimensional parametric nonlinear Schrödinger equation

$$\begin{cases} iu_t(t,x,\epsilon) = -u_{xx}(t,x,\epsilon) - \epsilon|u(t,x,\epsilon)|^2 u(t,x,\epsilon), \\ u(0,x) = u_0(x), \end{cases} \tag{4.68}$$

where u is a complex valued wave function, i is the imaginary unit, $|\cdot|$ is the modulus operator and ϵ is a parameter that belongs to the interval $\Gamma = [0.9, 1.1]$. We consider periodic boundary conditions, i.e., x belongs to a one-dimensional torus of length L. We consider the initial condition

$$u_0(x) = \frac{\sqrt{2}}{\cosh(x - x_0)} \exp(i\frac{c(x - x_0)}{2}), \tag{4.69}$$

for a positive constant c. In quantum mechanics, the quantity $|u(t,x)|^2$ represents the probability of finding the system in state x at time t. For the choice of $\epsilon = 1$, $|u(x,t)|$ becomes a solitary wave, and the initial condition will be transported in the positive x direction with a constant speed. For other choices of ϵ, the solution comprises an ensemble of solitary waves, moving in either direction [34].

By introducing the real and imaginary variables $u = p + iq$, we can rewrite (4.68) in canonical form as

$$\begin{cases} q_t = p_{xx} + \epsilon(q^2 + p^2)p, \\ p_t = -q_{xx} - \epsilon(q^2 + p^2)q, \end{cases} \tag{4.70}$$

with the Hamiltonian function

$$H(q,p) = \int_0^L (q_x^2 + p_x^2) + \frac{\epsilon}{2}(q^2 + p^2)^2 \, dx. \tag{4.71}$$

We discretize the torus into N equidistant points and take $\Delta x = L/N$, $x_i = i\Delta x$, $q_i = q(t, x_i, \epsilon)$ and $p_i = p(t, x_i, \omega)$ for $i = 1, \ldots, N$. A central finite differences scheme is used to discretize (4.70) as

$$\frac{d}{dt} z = \mathbb{J}_{2N} L z + \mathbb{J}_{2N} g(z). \tag{4.72}$$

Here $z = (q_1, \ldots, q_N, p_1, \ldots, p_n)^T$ and

$$L = \begin{pmatrix} D_{xx} & 0_N \\ 0_N & D_{xx} \end{pmatrix}. \tag{4.73}$$

Here g is a vector valued nonlinear function defined as

$$g(z) = \begin{pmatrix} (q_1^2 + p_1^2)q_1 \\ \vdots \\ (q_N^2 + p_N^2)q_N \\ (q_1^2 + p_1^2)p_1 \\ \vdots \\ (q_N^2 + p_N^2)p_N \end{pmatrix}. \tag{4.74}$$

We discretize the Hamiltonian to obtain

$$H_{\Delta x}(z) = \Delta x \sum_{i=1}^{N} \left(\frac{q_i q_{i-1} - q_i^2}{\Delta x^2} + \frac{p_i p_{i-1} - p_i^2}{\Delta x^2} + \frac{\epsilon}{4}(p_i^2 + q_i^2)^2 \right), \tag{4.75}$$

and use a Störmer-Verlet (2.36) scheme for time integration. Since the Hamiltonian function (4.75) is non-separable, this scheme is implicit. Hence, in each time iteration,

a system of nonlinear equations is solved using Newton's iteration. We summarize the physical and numerical parameters for the full model in the following table

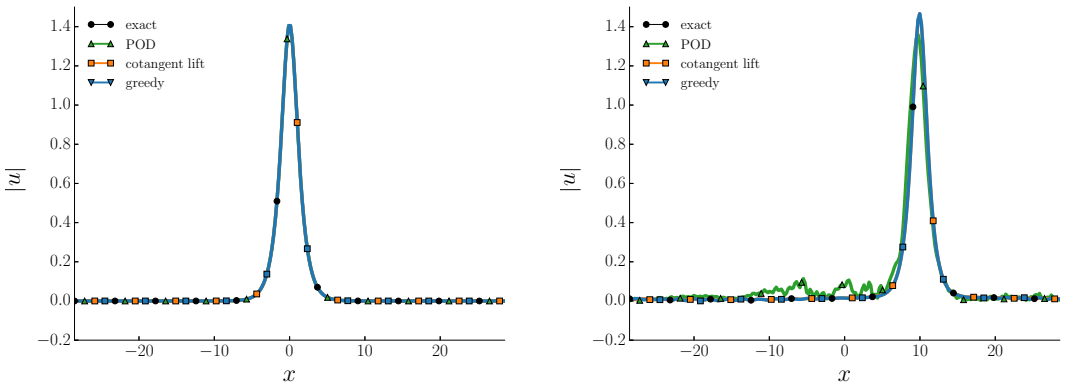
Domain length	$L = 2\pi/l$
Domain scaling factor	$l = 0.11$
wave speed	$c = 1$
No. grid points	$N = 256$
Space discretization size	$\Delta x = 0.2231$
Time discretization size	$\Delta t = 0.01$

Regarding computation of the nonlinear terms of reduced systems, we compare the DEIM with the symplectic DEIM. For generation of the DEIM we apply Algorithm 3.5 to the set of nonlinear snapshots. The method discussed in Section 4.5 is used to construct a reduced basis appropriate for the symplectic DEIM. The tolerance for the projection error is set to $\delta = 10^{-4}$.

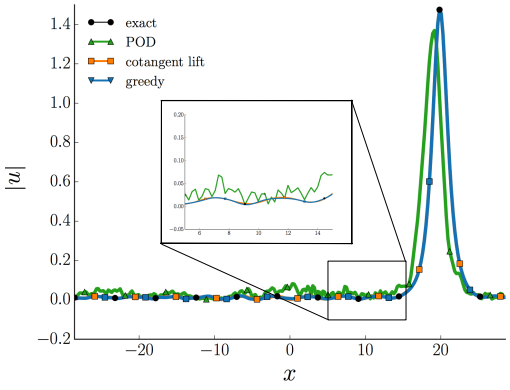
We compare the reduced system, obtained using the greedy algorithm, with that obtained by the cotangent lift, the complex SVD, DEIM, the symplectic DEIM and also POD. For the SVD-based methods, we discretize the parameter space $[0.9, 1.1]$ into $M = 500$ equidistant grid points across the discrete parameter space $\Gamma_M = \{\epsilon_1, \ldots, \epsilon_M\}$, and recover trajectory snapshots for each ϵ_i for $i = 1, \ldots, M$ in the snapshots matrix S. All reduced systems are taken to have identical sizes ($k = 90$ for the symplectic methods and $k = 180$ for the POD method). Following Algorithm 4.1 we construct the reduced system using the same discrete parameter space Γ_M. The tolerance for the error in the Hamiltonian is set to $\delta = 10^{-3}$. Moreover, for DEIM and symplectic DEIM, we construct bases of size $k' = 80$. Note that the reduced system, generated in the symplectic DEIM, will be of size $k + k' = 170$.

The cost of the offline stage is reduced by 80% when using the greedy method for constructing a symplectic basis of size $k = 90$, as compared to the SVD-based methods. The online stage, i.e., time integration for a new parameter in Γ, is generally more than 3 times faster than for the original system. We point out that the efficiency of the reduced systems are implementation and platform dependent and we expect further reduction as the size of the problem increases.

Figure 4.3 shows the solution of the Schrödinger equation for parameter value $\epsilon = 1.0932$ at $t = 0$, $t = 10$ and $t = 20$. We first compare the reduced system obtained by the greedy algorithm with the POD, the cotangent lift, and the complex SVD method. The size of the reduced systems are taken identical for all methods ($k = 180$ for POD and $k = 90$ for the rest). Although the decay of the singular values in Figure 4.5.(b)

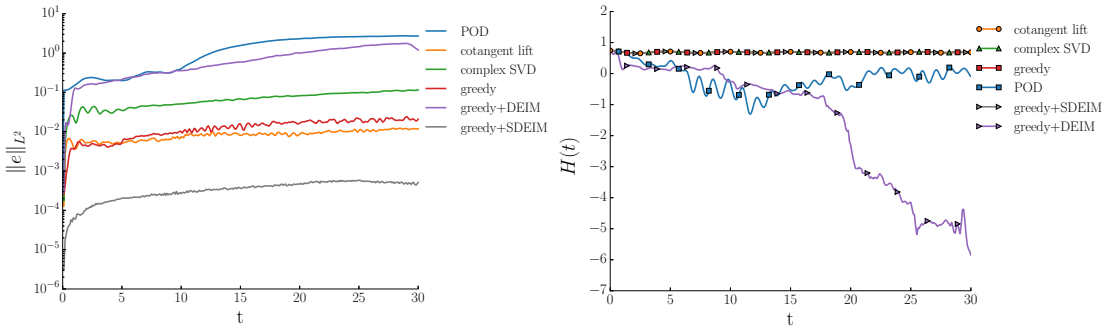


(a) $t = 0$ (b) $t = 10$



(c) $t = 20$

Figure 4.3 – The solution $|u(t,x)| = \sqrt{q^2 + p^2}$ at $t = 0$, $t = 10$ and $t = 20$ of the Nonlinear Schrödinger equation for parameter value $\epsilon = 1.0932$. Here the solution of the full system, together with the solution of the POD, cotangent lift, complex SVD and the greedy reduced system, is shown.



(a) (b)

Figure 4.4 – (a) Plot of the Hamiltonian function for $t \in [0, 30]$. (b) The L^2-error between the solution of the full system and the reduced system for different model reduction methods for $t \in [0, 30]$.

suggests that the accuracy of the POD reduced system should be comparable to that of the other methods, we observe signs of instabilities in the solution at $t = 10$. The greedy, the cotangent lift and the complex SVD method, on the other hand, generate a stable reduced system that accurately approximates the solution of the full model.

In Figure 4.4.(b) we observe that the symplectic methods preserve the Hamiltonian function, unlike the POD and the DEIM methods. We emphasise that the use of the reduced basis, obtained by the greedy, with the DEIM (purple line) does not preserve the symplectic structure as suggested in this figure.

Figure 4.4.(c) illustrates the L^2-error between the solution of the full model with the reduced systems, generated by different methods. We first observe that symplectic methods yield a lower computational error when compared to non-symplectic methods. Secondly, we observe that although the reduced systems from the cotangent lift and the complex SVD are of the same size, their accuracy differs by an order of magnitude. We notice that the greedy algorithm is slightly less accurate than the cotangent lift method while its offline computational cost is reduced by 80% when compared to the cotangent lift. Lastly we notice that the combination of the greedy reduced basis and DEIM yields large errors in the solution while the solution using the symplectic DEIM is very accurate. The symplectic DEIM is even more accurate than the greedy itself since it has been enriched by the nonlinear snapshots.

4.6.3 Numerical Convergence

In this section we discuss the numerical convergence of the symplectic greedy method introduced in Section 4.3. The exponential convergence properties of the conventional greedy was discussed in Section 3.3. Theorem 4.5 suggests that the symplectic greedy method has similar properties. To illustrate this we compare the convergence of the

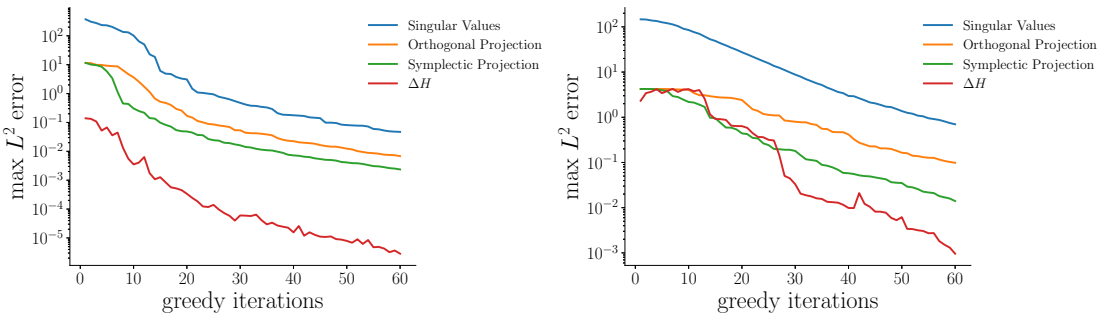


Figure 4.5 – (a) Convergence of the greedy method for the wave equation. (b) Convergence of the greedy method for the nonlinear Schrödinger equation equation.

conventional greedy with the convergence of the symplectic greedy method through the numerical simulations in Sections 4.6.1 and 4.6.2.

The decay of the singular values of the snapshot matrix for the parametric wave equation and the nonlinear Schrödinger equation are given in Figure 4.5. The decay rate of the singular values is a strong indicator for the decay rate of the Kolmogorov n-width of the solution manifold. We expect that the conventional greedy method and the symplectic greedy method provide a similar rate in the decay of the error.

Figure 4.5 shows the maximum L^2-error between the original system and the reduced system at each iteration of the different greedy methods. In this figure we find the conventional greedy with the orthogonal projection error as a basis selection criterion (orange), the symplectic greedy method with a symplectic projection error as a basis selection criterion (green), and the symplectic greedy method with energy loss ΔH as a basis selection criterion (red).

We observe that the decay rate of the error for the greedy method with the orthogonal projection and the greedy with the symplectic projection is similar to the decay of the singular values. This matches our expectation from Theorem 4.5. We also notice that the greedy method with the loss in Hamiltonian provides an excellent error indication as a basis selection criterion. However, as discussed in Section 4.4, the error in the Hamiltonian can sometimes fail to identify the best possible snapshot, e.g., when the initial condition is included in the reduced basis.

4.7 Conclusion

In this chapter, we have presented a greedy approach for the construction of a reduced system that preserves the geometric structure of Hamiltonian systems. An iteration of the greedy method comprises searching the parameter space using an error esti-

mator,e.g. the error in the Hamiltonian, to find the best basis vectors that increase the overall accuracy of the reduced basis. We argue that for a compact subset with exponentially small Kolmogorov n-width we recover exponentially fast convergence of the greedy algorithm. For the fast approximation of nonlinear terms, the basis obtained by the greedy was combined with a symplectic DEIM to construct a reduced system with a Hamiltonian that is arbitrary close to the Hamiltonian of the original system.

The numerical results demonstrate that the greedy method can save substantial computational cost in the offline stage as compared to alternative SVD-based techniques. Furthermore, since the reduced system obtained by the greedy method is Hamiltonian, the greedy method yields a stable reduced system. The symplectic DEIM effectively reduces the computational cost of approximate evaluation of nonlinear terms while preserving the stability and structure. Hence, the greedy method is an efficient model reduction technique that provides an accurate and stable reduced system for large-scale parametric Hamiltonian problems.

5 Symplectic Model Order Reduction With a Weighted Inner Product

In the previous chapters we discussed how MOR methods can substantially reduce the computational complexity of the problem by constructing a reduced configuration space. Exploration of the reduced space is then possible with significant acceleration [52, 48].

During the past decade, RB methods have demonstrated substantial lowering of the computational costs of solving elliptic and parabolic differential equations [55, 57]. However, as seen in Chapter 4, Development of MOR for hyperbolic problems remains a challenge. Such problems often arise from a set of conservation laws and invariants and this intrinsic structure is lost during MOR, resulting in a qualitatively wrong, and sometimes unstable reduced system [3].

Recently, the construction of RB methods that conserve intrinsic structures has attracted attention [2, 60, 36, 10, 24, 9, 81]. Structure preservation in MOR not only results in a physically meaningful reduced system, but can also enhance the robustness and stability of the reduced system. In system theory, conservation of passivity can be found in the work of [84, 46]. Energy preserving and inf-sup stable methods for finite element methods (FEM) are developed in [36, 6]. Also, a conservative MOR technique for finite-volume methods is proposed in [21].

Moreover, the simulation of reduced models incurs solution errors and the estimation of this error is essential in applications of MOR [50, 94, 37]. Finding tight error bounds for a general reduced system has shown to be computationally expensive and often impractical. Therefore, when one is interested in a cheap surrogate for the error or when the conserved quantity is an output of the system, it becomes imperative to preserve system structures of the reduced model.

In the context of Lagrangian and Hamiltonian systems, recent work provides a promising approach to the construction of robust and stable reduced systems. Carlberg, Tuminaro, and Boggs [24] suggest that a reduced order model of a Lagrangian system

be identified by an approximate Lagrangian on a reduced order configuration space. This allows the reduced system to inherit the geometric structure of the original system. A similar approach has been adopted in the work of Peng and Mohseni [81] and in the method discussed in Chapter 4 for Hamiltonian systems. They construct a low-order symplectic linear vector space, i.e. a vector space equipped with a symplectic 2-form, as the reduced space. Once the symplectic reduced space is generated, a symplectic projection results in a physically meaningful reduced system. A suitable time-stepping scheme then ensures preservation of the Hamiltonian structure of the reduced system. It is shown in [2, 81] that this approach preserves the overall dynamics of the original system and enhances the stability of the reduced system. Despite the success of these methods for MOR of Hamiltonian systems, these techniques are only compatible with the Euclidean inner product. Therefore, the computational structures that arise from a natural inner product of a problem will be lost during MOR.

Weak formulations and inner-products, defined on a Hilbert space, are at the core of the error analysis of many numerical methods for solving partial differential equations. Therefore, it is natural to seek MOR methods that consider such features. At the discrete level, these features often require a Euclidean vector space to be equipped with a generalized inner product, associated with a weight matrix X. Many works enabled conventional MOR techniques to be compatible with such inner products [97]. However, a MOR method that simultaneously preserves the symplectic symmetry of Hamiltonian systems remains unknown.

In this chapter, we seek to combine a classical MOR method, derived with respect to a weight matrix, with the symplectic MOR. A reduced system is constructed by orthogonally project a generalized Hamiltonian system onto the reduced space, with respect to a weighted inner product. The reduced system, however, carries the Hamiltonian structure and also the symplectic symmetry. It is shown that the new method can be viewed as the natural extension of to the one discussed in Chapter 4, and therefore retains the structure preserving features, e.g. symplecticity and stability. We also present a greedy approach for the construction of a generalized symplectic basis for the reduced system. Structured matrices are in general not norm bounded [61]. However, we show that the condition number of the basis generated by the greedy method is bounded by the condition number of the weight matrix X. Finally, to accelerate the evaluation of nonlinear terms in the reduced system, we present a variation of the discrete empirical interpolation method (DEIM) that preserves the symplectic structure of the reduced system.

The main contributions of this chapter are presenting a symplectic MOR technique that minimizes the model reduction error with respect to a general norm associated with a positive definite weight matrix X. A greedy method is proposed for the construction of a reduced basis that is both symplectic and ortho-normal with respect to the X-norm. Furthermore, the DEIM to efficiently evaluate nonlinear terms while preserving the

symplectic structure of the reduced system.

5.1 Generalization of the Symplectic Galerkin Projection

The error analysis of methods for solving partial differential equations often requires the use of a weighted inner product. This is particularly important when dealing with Hamiltonian systems, where the system energy induces a norm that is fundamental to the dynamics of the system. Furthermore, as we discussed in Section 3.2, the projection operator associated to an RB method is closely related to the inner-product defined on the high-fidelity space. We saw how a Galerkin projection operator can be constructed using a general inner-product in Section 3.2.2. However, the symplectic Galerkin projection introduced in Chapter 4 is only compatible with the Euclidean inner-product.

Recall the formulation of a Hamiltonian system defined on a $2n$-dimensional symplectic linear vector space $(\mathcal{Z}, \Omega)$ with a canonical basis $Z = \{e_i, f_i\}_{i=1}^n$

$$\begin{cases} \dfrac{d}{dt} z = \mathbb{J}_{2n} \nabla_z H(z), \\ z(0) = z_0. \end{cases} \tag{5.1}$$

Here, $H : \mathcal{Z} \to \mathbb{R}$ is the Hamiltonian and $z \in \mathcal{Z}$ is the state vector. Suppose that $\mathcal{Z}$ is equipped with an inner-product $< \cdot, \cdot >_X : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ such that $< a, b >_X = a^T X b$, for all $a, b \in \mathcal{Z}$ and some symmetric and positive-definite matrix X. Z is not in general orthonormal with respect to $<, >_X$. As the matrix form of Ω is $\mathbb{J}_{2n}$, Corollary 2.13 implies that Z is orthonormal with respect to the Euclidean inner product. Note that $z \in \mathcal{Z}$ implies that we can find $\alpha_i, \alpha_i' \in \mathbb{R}$, for $i = 1, \ldots, n$, such that $z = \alpha_i e_i + \alpha_i' f_i$. It yields

$$z = \alpha_i e_i + \alpha_i' f_i = \alpha_i X^{1/2} X^{-1/2} e_i + X^{1/2} X^{-1/2} \alpha_i' f_i. \tag{5.2}$$

By defining basis vectors $\tilde{e}_i = X^{-1/2} e_i$, $\tilde{e}_i = X^{-1/2} f_i$, and $\tilde{Z} = \{\tilde{e}_i, \tilde{f}_i\}_{i=1}^n$ with $\tilde{z}$ a state of a vector in this basis, we recover

$$z = X^{1/2} \left(\alpha_i \tilde{e}_i + \alpha_i' \tilde{f}_i \right) = X^{1/2} \tilde{z}. \tag{5.3}$$

We verify that $\tilde{Z}$ forms an orthonormal basis with respect to $< \cdot, \cdot >_X$

$$< \tilde{e}_i, \tilde{f}_j >_X = \tilde{e}_i^T X \tilde{f}_j = e_i^T X^{-1/2} X X^{-1/2} f_j = e_i^T f_j = 0. \tag{5.4}$$

Here we used the orthonormality of Z. Similarly we can show that $< \tilde{e}_i, \tilde{e}_j >_X = \delta_{ij}$ and $< \tilde{f}_i, \tilde{f}_j >_X = \delta_{ij}$, for $i = 1, \ldots, n$. Furthermore, we can verify that $\tilde{Z}$ is a symplectic

basis with respect to the symplectic form $\Omega_{J_{2n}}(a,b) = a^T J_{2n} b$, with $J_{2n} = X^{1/2}\mathbb{J}_{2n}X^{1/2}$

$$\Omega_{J_{2n}}(\tilde{e}_i, \tilde{f}_j) = \tilde{e}_i^T J_{2n}\tilde{f}_j = e_i^T X^{-1/2}X^{1/2}\mathbb{J}_{2n}X^{1/2}X^{-1/2}f_j = e_i^T\mathbb{J}_{2n}f_j = \delta_{ij}. \tag{5.5}$$

Here we used symplecticity [1] of Z with respect to $\mathbb{J}_{2n}$. Similarly we can verify that $\Omega_{J_{2n}}(\tilde{e}_i,\tilde{e}_j) = \Omega_{J_{2n}}(\tilde{f}_i,\tilde{f}_j) = 0$, for $i = 1,\dots,n$. Using the state transformation $z = X^{1/2}\tilde{z}$, the Hamiltonian system (5.1) takes the form [2]

$$\begin{cases} \dfrac{d}{dt}\tilde{z} = J_{2n}^{-1}\nabla_{\tilde{z}}H_X(\tilde{z}), \\ \tilde{z}(0) = X^{-1/2}z_0. \end{cases} \tag{5.6}$$

where $H_X = -H(X^{1/2}\tilde{z})$.

Definition 5.1. *A matrix $\tilde{A} \in \mathbb{R}^{2n\times 2k}$ is called* J_{2n}-symplectic, *if it transforms J_{2n} into the standard symplectic matrix $\mathbb{J}_{2k}$, i.e.,*

$$\tilde{A}^T J_{2n}\tilde{A} = \mathbb{J}_{2k}. \tag{5.7}$$

Definition 5.2. *The symplectic inverse of a J_{2n}-symplectic matrix $\tilde{A} \in \mathbb{R}^{2n\times 2k}$ is defined as*

$$\tilde{A}^+ := \mathbb{J}_{2k}^T\tilde{A}^T J_{2n}. \tag{5.8}$$

Note that since this definition is an extension of the symplectic inverse defined in Section 4.1, we may use the "+" superscript for both. The following theorem summarizes the properties of the symplectic inverse in this generalized setting.

Proposition 5.1. *Let $\tilde{A} \in \mathbb{R}^{2n\times 2k}$ be a J_{2n}-symplectic basis where $J_{2n} \in \mathbb{R}^{2n\times 2n}$ is a full rank and skew-symmetric matrix. Furthermore, suppose that $\tilde{A}^+ = \mathbb{J}_{2k}^T\tilde{A}^T J_{2n}$ is the symplectic inverse. Then the following holds:*

(a) $\tilde{A}^+\tilde{A} = I_{2k}$.

(b) $(\tilde{A}^+)^T$ *is* J_{2n}^{-1}*-symplectic.*

(c) $\left(\left(\left(\tilde{A}^+\right)^T\right)^+\right)^T = \tilde{A}$.

[1] $\Omega_{\mathbb{J}_{2n}}(e_i,f_i) = e_i^T\mathbb{J}_{2n}f_j$

[2] substituting $z = X^{1/2}\tilde{z}$ in (5.1) yields

$$\frac{d}{dt}X^{1/2}\tilde{z} = \mathbb{J}_{2n}\nabla_z H(X^{1/2}\tilde{z}) = -\mathbb{J}_{2n}^{-1}X^{-1/2}\nabla_{\tilde{z}}H(X^{1/2}\tilde{z})$$
$$\implies \frac{d}{dt}\tilde{z} = X^{-1/2}\mathbb{J}_{2n}^{-1}X^{-1/2}\nabla_{\tilde{z}}(-H(X^{1/2}\tilde{z}))$$

(d) Let $J_{2n} = X^{1/2}\mathbb{J}_{2n}X^{1/2}$. Then $\tilde{A}$ is ortho-normal with respect to the $<,>_X$, if and only if $(\tilde{A}^+)^T$ is ortho-normal with respect to the $<,>_{X^{-1}}$.

Proof. It is straightforward to show all statements using the definition of a symplectic basis. □

Note statement (d) in Proposition 4.1 does not hold in the generalized setting, i.e., when $\tilde{A}$ is orthonormal with respect to $<,>_X$, $(\tilde{A}^+)^T \neq \tilde{A}$. This is particularly important when constructing a basis for nonlinear terms in Section 5.5.

In this chapter, we indicate a non-standard symplectic matrix/transformation/subspace with "~" overscript. We are now ready to motivate the choice of the basis $\tilde{Z}$ in (5.6).

Lemma 5.2. *A full rank J_{2n}-symplectic linear transformation $\tilde{A} \in \mathbb{R}^{2n\times 2n}$ transforms* (5.6) *into the standard Hamiltonian form.*

Proof. Let $\tilde{A} \in \mathbb{R}^{2n\times 2n}$ be a J_{2n}-symplectic mapping. We define the state transformation $\tilde{z} = \tilde{A}y$. Note that since $\tilde{A}$ is a square matrix, we can indeed require this relation to be an equality. This transforms (5.6) into

$$\frac{d}{dt}y = \tilde{A}^+ J_{2n}^{-1}(\tilde{A}^+)^T \nabla_y H_X(\tilde{A}y). \tag{5.9}$$

However, **??** indicates that $(\tilde{A}^+)^T$ is J_{2n}^{-1}-symplectic, thus,

$$\frac{d}{dt}y = \mathbb{J}_{2n}\nabla_y H_X(\tilde{A}y). \tag{5.10}$$

□

Note that even though (5.1) and (5.10) are both in the standard form, they are not identical. Furthermore, the form of (5.6) is preferred from the MOR standpoint, since an orthonormal basis with respect to $<,>_X$ can be constructed such that it preserves the Hamiltonian form (Lemma 5.2).

Suppose that a $2k$-dimensional linear vector space $\tilde{\mathcal{A}}_{2k}$, with $k \ll n$, is provided such that it approximates well the solution manifold $\mathcal{M}_H$ of (5.6). Let $\tilde{E} = \{\tilde{e}_i, \tilde{f}_i\}_{i=1}^k$ be the basis for this subspace and construct the matrix

$$\tilde{A}_{2k} = [\tilde{e}_1, \ldots, \tilde{e}_k, \tilde{f}_1, \ldots, \tilde{f}_k] \in \mathbb{R}^{2n\times 2k}. \tag{5.11}$$

We require $\tilde{A}_{2k}$ be a J_{2n}-symplectic basis and approximate a solution to (5.6) as $\tilde{z} \approx$

$\tilde{A}_{2k}y$ to write

$$\tilde{A}_{2k}\frac{d}{dt}y = J_{2n}^{-1}(\tilde{A}_{2k}^{+})^{T}\nabla_y H_X(\tilde{A}_{2k}y) + J_{2n}^{-1}r(z). \tag{5.12}$$

Assuming that the error vector r is symplectically orthogonal to $\tilde{A}_{2k}$ and using Lemma 5.2, we recover

$$\begin{cases} \dfrac{d}{dt}y = \mathbb{J}_{2k}\nabla_y H_X(\tilde{A}_{2k}y), \\ y(0) = \tilde{A}_{2k}^{+}\tilde{z}_0. \end{cases} \tag{5.13}$$

The projection operator that projects members of $\tilde{Z}$ onto $\tilde{\mathcal{A}}_{2k}$ is the *generalized symplectic Galerkin projection* and is defined as

$$P^{\text{symp}}_{X,\tilde{A}_{2k}}(\tilde{z}) = \tilde{A}_{2k}\tilde{A}_{2k}^{+}\tilde{z}. \tag{5.14}$$

Finally, as the final goal is to approximate z, the solution to (5.1), we write

$$\begin{cases} \dfrac{d}{dt}y = \mathbb{J}_{2k}\nabla_y \tilde{H}(y), \\ y(0) = \mathbb{J}_{2n}^{T}A^{T}X\mathbb{J}_{2n}z_0. \end{cases} \tag{5.15}$$

Where $A_{2k} = X^{-1/2}\tilde{A}_{2k}$ and $\tilde{H}(y) = -H(XA_{2k}y)$ is the reduced Hamiltonian. Accordingly, the projection operator $P^{\text{symp}}_{X,A_{2k}} : \mathcal{Z} \to \mathcal{A}_{2k}$ can be written as

$$P^{\text{symp}}_{X,A_{2k}}(z) = X^{-1/2}\tilde{A}_{2k}\tilde{A}_{2k}^{+}X^{1/2}z = A_{2k}\mathbb{J}_{2k}^{T}A_{2k}^{T}X\mathbb{J}_{2n}Xz. \tag{5.16}$$

We can check that $P^{\text{symp}}_{X,A_{2k}}$ is indeed a projection operator

$$P^{\text{symp}}_{X,A_{2k}} \circ P^{\text{symp}}_{X,A_{2k}} = A_{2k}\underbrace{\mathbb{J}_{2k}^{T}A_{2k}^{T}X\mathbb{J}_{2n}XA_{2k}}_{=\tilde{A}^{+}\tilde{A}=I_{2k}}\mathbb{J}_{2k}^{T}A_{2k}^{T}X\mathbb{J}_{2n}X = P^{\text{symp}}_{X,A_{2k}} \tag{5.17}$$

Sections 5.2 and 5.4 discuss how to efficiently construct the reduced basis A_{2k}.

5.2 Proper Symplectic Decomposition Revisited

Let S be the snapshot matrix of the Hamiltonian system (5.6). Similar to the approach presented in Sections 3.2 and 4.2, we seek to minimize the projection error with respect to the $P^{\text{symp}}_{X,A}$, defined in (5.16), and the X-norm, i.e., finding the solution to the minimization

$$\begin{aligned} &\underset{A\in\mathbb{R}^{2n\times 2k},s\in S}{\text{minimize}} && \sum_{s\in S}\|s - P^{\text{symp}}_{X,A}(s)\|_X^2, \\ &\text{subject to} && \mathbb{J}_{2k}^{T}A^{T}X\mathbb{J}_{2n}XA = I_{2k}. \end{aligned} \tag{5.18}$$

Here, the constraint ensures that $P^{\text{symp}}_{X,A}$ is a projection operator, see (5.17). It follows

$$\begin{aligned}\sum_{s\in S}\|s-P^{\text{symp}}_{X,A}(s)\|_X^2 &= \sum_{s\in S}\|s-A\mathbb{J}_{2k}^T A^T X\mathbb{J}_{2n}Xs\|_X^2\\ &= \sum_{s\in S}\|X^{1/2}s-X^{1/2}A\mathbb{J}_{2k}^T A^T X\mathbb{J}_{2n}Xs\|_2^2\\ &= \|X^{1/2}S-X^{1/2}A\mathbb{J}_{2k}^T A^T X\mathbb{J}_{2n}XS\|_F^2\\ &= \|\tilde{S}-\tilde{A}\tilde{A}^+\tilde{S}\|_F^2.\end{aligned} \tag{5.19}$$

Here $\tilde{S}=X^{1/2}S$, $\tilde{A}=X^{1/2}A$ and $\tilde{A}^+=\mathbb{J}_{2k}^T\tilde{A}^T J_{2n}$ is the symplectic inverse of $\tilde{A}$ with respect to the skew-symmetric matrix $J_{2n}=X^{1/2}\mathbb{J}_{2n}X^{1/2}$, introduced in Section 5.1. With this notation, the constraint in (5.18) can be reformulated as $\tilde{A}^+\tilde{A}=I_{2k}$ which is equivalent to $\tilde{A}^T J_{2n}\tilde{A}=\mathbb{J}_{2k}$. In other words, this condition implies that $\tilde{A}$ has to be a J_{2n}-symplectic matrix. Finally we can rewrite the minimization (5.18) as

$$\begin{aligned}&\underset{\tilde{A}\in\mathbb{R}^{2n\times 2k}}{\text{minimize}} \quad \|\tilde{S}-P^{\text{symp}}_{X,\tilde{A}}(\tilde{S})\|_F,\\ &\text{subject to} \quad \tilde{A}^T J_{2n}\tilde{A}=\mathbb{J}_{2k}.\end{aligned} \tag{5.20}$$

where $P^{\text{symp}}_{X,\tilde{A}}=\tilde{A}\tilde{A}^+$ is the symplectic projection with respect to the X-norm onto $\mathcal{A}$, the column span of A. At first glance, the minimization (5.20) looks similar to (4.15). However, since $\tilde{A}$ is J_{2n}-symplectic, and the projection operator depends on X, we need to seek an alternative approach to find a near optimal solution to (5.20).

Direct approaches for solving (5.20) are impractical. Furthermore, there are no known SVD-type methods to solve (5.20). However, the greedy generation of the symplectic basis can be generalized to generate a near optimal basis $\tilde{A}$. The generalized greedy method is discussed in Section 5.4.

5.3 Stability Conservation

We discussed in Section 4.1 that a Hamiltonian reduced system, constructed by the projection $P^{\text{symp}}_{I,A}$, preserves the stability of stable equilibrium points of (4.10), and therefore, preserves the overall dynamics. In this section, we establish that the stability of the equilibrium points is also conserved using the projection operator $P^{\text{symp}}_{X,\tilde{A}}$.

Theorem 5.3. *[2] Consider a Hamiltonian system of the form* (5.1) *with a Hamiltonian $H\in C^2$ together with the reduced system* (5.15)*. Suppose that z_e is a strict local minimum of H. Furthermore, suppose that H (or $-H$) is a Lyapunov function satisfying Proposition 4.3. If we can find an open ball neighborhood S of z_e such that Range$(XA)\cap S\neq\emptyset$, then the reduced system* (5.15) *has a stable equilibrium point in Range$(XA)\cap S$.*

Proof. Assume that H satisfies the conditions in Proposition 4.3 for a Lyapunov function. Since z_e is a local minimum of H, smoothness of H implies that $\nabla_z H(z_e) = 0$, and threfore z_e is a Lyapunov stable point for (5.1). Furthermore, since z_e is a strict local minimum, we can find an open ball S of z_e such that $H(z_e) < H(z)$, for any $z \in \overline{S}$ and $z \neq z_e$, where $\overline{S}$ is the closure of B. Note that $\overline{S}$ is bounded. Since $H \in C^2$ and $\nabla^2 H$ is positive definite at z_e, we also take S to be small enough such that $\nabla^2 H > 0$, for all $z \in S$. Define $c := \inf_{z \in \partial S} H(z)$, where ∂S is the boundary of S. Since H is continues and z_e is a strict minimum we can assume that $H(z) < c$, for all $z \in S$ [3].

Let $S_A = \text{Range}(XA) \cap S$. Since $\text{Range}(XA)$ is a linear vector space, then S_A is an open set. Furthermore, for any $z \in S_A$, $H(z) < c$.

We now show that $H|_{S_A}$ attains its minimum inside S_A. Let $c_{\min} = \inf_{z \in S_A} H(z)$. c_{min} exists since H has a minimum on S. We can find a sequence $\{H(z_i)\}_{i=1}^{\infty}$, with $z_i \in S_A$, such that $H(z_i) \to c_{\min} < c$. This implies that $z_i \to z_0$, for some $z_0 \in \overline{S_A}$, since $H \in C^2$. Note that $\overline{S_A}$ is bounded since S is bounded. However, z_0 does not belong to ∂S_A since $\inf_{z \in \partial S} W(z) = c > c_{min}$. Therefore $z_0 \in S_A$.

We claim that $y_e = \mathbb{J}_{2k}^T A^T X \mathbb{J}_{2n} z_0$ is a stable equilibrium point for the reduced system (5.15). Let $\tilde{W}(y) = -\tilde{H}(y) = H(XAy)$. Note that $\tilde{W}$ attains its local minimum at y_e. Furthermore, $\nabla \tilde{W}(y_e) = 0$. Also we have

$$\nabla^2 \tilde{W} = A^T X \nabla^2 H X A \tag{5.21}$$

is a positive definite matrix. Finally, since the reduced system is a Hamiltonian system, Corollary 2.11 implies that any trajectory φ_t of (5.15) satisfies $\frac{d}{dt}\tilde{W}(\varphi_t) = 0$. Therefore $\tilde{W}$ is a Lyapunov function for (5.15) and y_e is a stable equilibrium point for (5.15), in the Lyapunov sense.

Similar derivations confirms the theorem when $-H$ is a candidate for a Lyapunov function. □

A reduced basis that is constructed accurate enough that satisfy the conditions in Theorem 5.3, guarantees to preserve the stability of the stable equilibrium points, and therefore, preserves the overall dynamics of the high-fidelity system.

5.4 Greedy Generation of a J_{2n}-Symplectic Basis

In this section we modify the greedy algorithm introduced in Section 4.3 to construct a J_{2n}-symplectic basis. We recall that ortho-normalization is an essential step in

[3] If there is no such open ball, we can construct a sequence $\{z\}_{i=1}^{\infty}$ such that $z_i \to z_e$ and $H(z_i) = c$, for all i, implying that $H(z_e) = c$ which is a contradiction.

greedy approaches to basis generation. Here, we summarize a variation of the GS orthogonalization process, known as the *symplectic GS* process.

Suppose that $\Omega_{J_{2n}}$ is a symplectic form defined on $\mathcal{Z} = \mathbb{R}^{2n}$ such that $\Omega_{J_{2n}}(x, y) = x^T J_{2n} y$, for all $x, y \in \mathbb{R}^{2n}$ and some full rank and skew-symmetric matrix $J_{2n} = X^{1/2}\mathbb{J}_{2n}X^{1/2}$. Let $\tilde{E}_{2k} = \{\tilde{e}_i, T^{-1}(\tilde{e}_i)\}_{i=1}^{k}$ be an ortho-symplectic basis with respect to $<,>_X$ and $\Omega_{J_{2n}}$, where the linear transformation T is defined in Proposition 2.12. It can be verified that $T^{-1}(z) = X^{-1/2}\mathbb{J}_{2n}^T X^{1/2} z$. Furthermore, let $\tilde{A}_{2k}$ be the matrix that contains these vectors in its columns and $\tilde{\mathcal{A}}_{2k}$ the space spanned by columns of $\tilde{A}_{2k}$. We seek to add $z \notin \tilde{\mathcal{A}}_{2k}$ to $\tilde{A}_{2k}$ to enhance the overall accuracy of the reduced basis measured in the L^∞ norm. We J_{2n}-orthogonalize z with respect to the basis vectors in $\tilde{E}_{2k}$, i.e., we construct the vector

$$\hat{z} = z - P^{\text{symp}}_{X,\tilde{\mathcal{A}}_{2k}}(z). \tag{5.22}$$

Let us introduce $\tilde{e}_{k+1} = \hat{z}/\|\hat{z}\|_X$. According to Proposition 2.12, the next pair of basis vectors $\{\tilde{e}_{k+1}, T^{-1}(\tilde{e}_{k+1})\}$ are ortho-symplectic to $\tilde{E}_{2k}$. Finally, the basis generated at the $(k+1)$-th step of the greedy method is $\tilde{E}_{2k+2} = \tilde{E}_{2k} \cup \{\tilde{e}_{k+1}, T^{-1}(\tilde{e}_{k+1})\}$ and the corresponding matrix is assembled as

$$\tilde{A}_{2k+2} = [\tilde{e}_1, \ldots, \tilde{e}_{k+1}, T^{-1}(\tilde{e}_1), \ldots, T^{-1}(\tilde{e}_{k+1})]. \tag{5.23}$$

We note that the symplectic GS orthogonalization process is chosen due to its simplicity. However, in problems where there is a need for a large basis, this process might be impractical. In such cases, one may use a backward stable routine, e.g. the isotropic Arnoldi method or the isotropic Lanczos method [72].

It is well known that a symplectic basis, in general, is not norm bounded [61]. The following theorem guarantees that the greedy method for generating a J_{2n}-symplectic basis yields a bounded basis.

Theorem 5.4. *The basis generated by the greedy method for constructing a J_{2n}-symplectic basis is orthonormal with respect to the X-norm.*

Proof. Let $\tilde{A}_{2k} = [\tilde{e}_1, \ldots, \tilde{e}_k, T^{-1}(\tilde{e}_1), \ldots, T^{-1}(\tilde{e}_k)]$ be the J_{2n}-symplectic basis generated at the kth step of the greedy method. Using the fact that $\tilde{A}_{2k}$ is J_{2n}-symplectic, one can check that

$$\langle \tilde{e}_i, \tilde{e}_j \rangle_X = \left\langle T^{-1}(\tilde{e}_i), T^{-1}(\tilde{e}_j) \right\rangle_X = \Omega_{J_{2n}}(\tilde{e}_i, T^{-1}(\tilde{e}_j)) = \delta_{i,j}, \quad i, j = 1, \ldots, k, \tag{5.24}$$

and

$$\left\langle \tilde{e}_i, T^{-1}(\tilde{e}_j) \right\rangle_X = \Omega_{J_{2n}}(\tilde{e}_i, \tilde{e}_j) = 0 \quad i, j = 1, \ldots, k, \tag{5.25}$$

where $\delta_{i,j}$ is the Kronecker delta function. This ensures that $\tilde{A}_{2k}^T X \tilde{A}_{2k} = I_{2k}$, i.e., $\tilde{A}_{2k}$ is an ortho-normal basis with respect to the X-norm. □

We note that if we take $X = I_{2n}$, then the greedy process generates a $\mathbb{J}_{2n}$- symplectic basis. With this choice, the greedy method discussed reduces to the greedy process discussed in Section 4.3. Therefore, the symplectic model reduction with a weight matrix X is a generalization of the method discussed in Chapter 4.

We notice that $X^{1/2}$ does not explicitly appear in the reduced Hamiltonian system (5.15). Therefore, it is desirable to compute $A_{2k} = X^{-1/2}\tilde{A}_{2k}$ without requiring the computation of the matrix square root of X. It is easily checked that the matrix $B_{2k} := X^{1/2}\tilde{A}_{2k} = XA_{2k}$ is $\mathbb{J}_{2n}$-symplectic and orthonormal. We can modify the J_{2n}-orthogonalization to obtain a $\mathbb{J}_{2n}$-orthogonalization, i.e., we seek $\alpha \in \mathbb{R}^{2k}$ such that

$$\Omega_{\mathbb{J}_{2n}}(w + B_{2k}\alpha, \bar{y}) = 0, \quad \forall \bar{y} \in \text{colspan}(B_{2k}), \tag{5.26}$$

where $w = X^{1/2}z$. From Chapter 4 we know that (5.26) has the unique solution $\alpha_i = -\Omega_{\mathbb{J}_{2n}}(z, \mathbb{J}_{2n}^T\hat{e}_i)$ for $i \leq k$ and $\alpha_i = \Omega_{\mathbb{J}_{2n}}(z, \hat{e}_i)$ for $i > k$, where $\hat{e}_i$ is the ith column vector of B_{2k}. Furthermore, we take

$$\hat{e}_{k+1} = \hat{z}/\|\hat{z}\|_2, \quad \hat{z} = w + B_{2k}\alpha, \tag{5.27}$$

and the next basis matrix is assembled as

$$B_{2k+2} = [\hat{e}_1, \ldots, \hat{e}_{k+1}, \mathbb{J}_{2n}^T\hat{e}_1, \ldots, \mathbb{J}_{2n}^T\hat{e}_{k+1}]. \tag{5.28}$$

One can recover e_{k+1} from the relation $e_{k+1} = X^{-1/2}\hat{e}_{k+1}$. However, since we are interested in the matrix A_{2k+2} and not $\tilde{A}_{2k+2}$, we can solve the system

$$XA_{2k+2} = B_{2k+2}, \tag{5.29}$$

for A_{2k+2}. This eliminates the need to compute of $X^{1/2}$.

To identify the best vectors to be added to a set of basis vectors, we may use error functions similar to those introduced in Chapter 4. The projection error can be used to identify the temporal snapshot that is worst approximated by a given basis $\tilde{A}_{2k}$:

$$z_{k+1} := \underset{z \in \{z(t_i)\}_{i=1}^{N_t}}{\arg\max} \|z - P_{X,A}^{\text{symp}}(z)\|_X, \tag{5.30}$$

where $P_{X,A}^{\text{symp}}$ is defined in (5.16). Alternatively we can use the loss in the Hamiltonian (4.13) for parameter dependent problems. We summarize the greedy method for generating a J_{2n}-symplectic matrix in Algorithm 5.1.

Algorithm 5.1 the generalized symplectic greedy for generating a symplectic reduced basis

Input: weight matrix X, parameter space $\mathbb{P}$, error indicator function η , symplectic reduced basis A_{2k}.

1: find $\mu^* := \arg\max_{\mu\in\mathbb{P}} \eta(\mu)$.
2: compute the temporal snapshots S_{t,μ^*}.
3: Find the snapshot with maximum projection error

$$z := \arg\max_{s\in S_{t,\mu_{k+1}}} \|s - P^{\text{symp}}_{X,A_{2k}}(s)\|_X.$$

4: compute $\hat{z} = Xz$ and $B_{2k} = XA_{2k}$.
5: apply $\mathbb{J}_{2n}$-orthonormalization on $\hat{z}$ to obtain $\hat{e}_{k+1}$.
6: solve $Xe_{k+1} = \hat{e}_{k+1}$ and $Xf_{k+1} = \mathbb{J}_{2n}^T\hat{e}_{k+1}$ for e_{k+1} and f_{k+1}.
7: assemble

$$A_{2k+2} = [e_1, \ldots, e_{k+1}, f_1, \ldots, f_{k+1}].$$

Output: symplectic reduced basis A_{2k+2}.

It was discussed in Section 4.4 that under natural assumptions on the solution manifold of a Hamiltonian system, the symplectic greedy method for symplectic basis generation converges exponentially fast. We expect the generalized greedy method, equipped with the error function (5.30), to converge as fast, since the X-norm is topologically equivalent to the standard Euclidean norm [44], for a full rank matrix X.

5.5 Efficient Evaluation of Nonlinear Terms

As discussed previously, the evaluation of the nonlinear term in (5.15) retains a computational complexity proportional to the size of the full order system (5.1). To overcome this, we take an approach similar to that in Section 4.5. Let $H(z) = H_1(z) + H_2(z)$ such that $\nabla_z H_1(z) = Lz$, for some $L \in \mathbb{R}^{2n\times 2n}$ and $\nabla_z H_2(z) = g(z)$, where $g : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is a vector valued function. We use a J_{2n}-symplectic reduced basis $\tilde{A}$ to construct a reduced Hamiltonian system from (5.6) as

$$\begin{aligned}\dot{y} &= \tilde{A}^+X^{-1/2}\mathbb{J}_{2n}\nabla_z H_1(z) + \tilde{A}^+X^{-1/2}\mathbb{J}_{2n}\nabla_z H_2(z)\\ &= \tilde{A}^+X^{-1/2}\mathbb{J}_{2n}X^{-1/2}(\tilde{A}^+)^T\nabla_y H_1(X^{1/2}\tilde{A}y) + \tilde{A}^+X^{-1/2}\mathbb{J}_{2n}\nabla_z H_2(z)\\ &= -\mathbb{J}_{2k}A^TXLXAy + \tilde{A}^+X^{-1/2}\mathbb{J}_{2n}g(z).\end{aligned} \tag{5.31}$$

Here we used $\nabla_y H_1(X^{1/2}\tilde{A}y) = A^TXLXAy$. The matrix $L_r := A^TXLXA$ can be computed in the offline phase. To accelerate the evaluation of the nonlinear term, we

apply the DEIM on $g(z)$ to obtain

$$\dot{y} = -\mathbb{J}_{2k} A^T X L X A y + \tilde{A}^+ X^{-1/2} \mathbb{J}_{2n} U (P^T U)^{-1} P^T g(X A y). \tag{5.32}$$

Here U is a basis for the nonlinear snapshots, and P is the interpolating index matrix (see Sections 3.5 and 4.5). For a general choice of U, the reduced system (5.32) does not maintain a Hamiltonian form. Note that $g(z) = \nabla_z H_2(z) = X^{-1/2}(\tilde{A}^+)^T \nabla_y H_2(X^{1/2}\tilde{A}y)$. Substituting this into (5.32) yields

$$\dot{y} = -\mathbb{J}_{2k} L_r y + \tilde{A}^+ X^{-1/2} \mathbb{J}_{2n} U (P^T U)^{-1} P^T X^{-1/2} (\tilde{A}^+)^T \nabla_y H_2(X^{1/2}\tilde{A}y). \tag{5.33}$$

Freedom in the choice of the basis U allows us to require $U = X^{-1/2}(\tilde{A}^+)^T$ which reduces the expression in (5.33) to

$$\dot{y} = -\mathbb{J}_{2k} L_r y - \mathbb{J}_{2k} \nabla_y H_2(X^{1/2}\tilde{A}y). \tag{5.34}$$

This is a Hamiltonian function identified by the reduced Hamiltonian $\tilde{H} = -\frac{1}{2} y^T L_r y - H_2(X^{1/2}\tilde{A}y)$. The reduced system yields

$$\begin{cases} \dot{y}(t) = -\mathbb{J}_{2k} L_r y - \mathbb{J}_{2k} (P^T \mathbb{J}_{2n}^T X A \mathbb{J}_{2k})^{-1} P^T g(z), \\ y(0) = \mathbb{J}_{2k}^T A^T X \mathbb{J}_{2n} z_0. \end{cases} \tag{5.35}$$

Let us now discuss how to ensure that $X^{-1/2}(\tilde{A}^+)^T$ is a basis for the nonlinear snapshots. Note that if $z \in \text{colspan}\left(X^{-1/2}(\tilde{A}^+)^T\right)$ then $X^{1/2} z \in \text{colspan}\left((\tilde{A}^+)^T\right)$. Therefore, it is sufficient to require $(\tilde{A}^+)^T$ to be a basis for $X^{1/2} S_{t,\mu}$, the nonlinear snapshots. Proposition 5.1 suggests that $(\tilde{A}^+)^T$ is a J_{2n}^{-1}-symplectic basis and that the transformation between $\tilde{A}$ and $(\tilde{A}^+)^T$ does not affect the symplectic feature of the bases. Consequently, from $\tilde{A}$ we may recover $(\tilde{A}^+)^T$ and enrich it with the nonlinear snapshots $\{X^{1/2} s\}_{s \in S_g}$. Once $(\tilde{A}^+)^T$ represents the nonlinear term with the desired accuracy, we may compute $\tilde{A} = \left(\left((\tilde{A}^+)^T\right)^+\right)^T$ to obtain the reduced basis for (5.35). Proposition 5.1 implies that $(\tilde{A}^+)^T$ is ortho-normal with respect to the X^{-1}-norm. This affects the ortho-normalization process. We note that greedy approaches to basis generation do not generally result in a minimal basis in the L^2 norm, but rather an optimal one in the L^∞ norm.

As discussed in Section 5.4 it is desirable to eliminate the computation of $X^{\pm 1/2}$. Having $z \in \text{colspan}\left(X^{-1/2}(\tilde{A}^+)^T\right)$ implies that $z \in \text{colspan}(\mathbb{J}_{2n}^T X A \mathbb{J}_{2k})$. Note that Algorithm 5.1 constructs a $\mathbb{J}_{2n}$-symplectic matrix XA and $\mathbb{J}_{2n}^T X A \mathbb{J}_{2k}$ is the symplectic inverse of XA with respect to the standard symplectic matrix $\mathbb{J}_{2n}$. Given e as a candidate for enriching $X^{-1/2}(\tilde{A}^+)^T$ we may instead enrich $\mathbb{J}_{2n}^T X A \mathbb{J}_{2k}$ with e. This process

eliminates the computation of $X^{\pm 1/2}$. We summarize the process of generating a basis for the nonlinear terms in Algorithm 5.2.

Algorithm 5.2 generation of a DEIM basis in the generalized setting

Input: reduced basis A from Algorithm 5.1, Nonlinear snapshots S_g, Tolerance δ

1: compute $B = XA$.
2: compute $(B^+)^T = \mathbb{J}_{2n}^T B \mathbb{J}_{2k} = [e_1, \ldots, e_k, \mathbb{J}_{2n}^T e_1, \ldots, \mathbb{J}_{2n}^T e_k]$.
3: **while** $\|s - P^{\text{symp}}_{I,(B^+)^T}(s)\|_2 > \delta$ for any $s \in S_g$ **do**
4: $\quad s^* := \arg\max_{s \in \mathcal{S}_g} \|s - P^{\text{symp}}_{I,(B^+)^T}(s)\|_2$.
5: $\quad \mathbb{J}_{2n}$-orthogonalize s^* to obtain e_{k+1}.
6: $\quad$ assemble $(B^+)^T = [e_1, \ldots, e_{k+1}, \mathbb{J}_{2n}^T e_1, \ldots, \mathbb{J}_{2n}^T e_{k+1}]$.
7: **end while**
8: set $XA = \left(\left(B^+)^T \right)^+ \right)^T$.

Output: $\mathbb{J}_{2n}$-symplectic basis XA.

5.6 Numerical Results

Let us now discuss the performance of the symplectic model reduction with a weighted inner product. In Sections 5.6.1 and 5.6.2 we apply the model reduction to equations of a vibrating elastic beam without and with a cavity, respectively. We examine the evaluation of the nonlinear terms in the model reduction of the sine-Gordon equation, in Section 5.6.3.

5.6.1 The Elastic Beam Equation

Consider the equations governing small deformations of a clamped elastic body $\Gamma \subset \mathbb{R}^3$ as

$$\begin{cases} u_{tt}(t,x) = \nabla \cdot \sigma + f, & x \in \Gamma, \\ u(0,x) = \vec{0}, & x \in \Gamma, \\ \sigma \cdot \hat{n} = \tau, & x \in \partial\Gamma_\tau, \\ u(t,x) = \vec{0}, & x \in \partial\Gamma \backslash \partial\Gamma_\tau, \end{cases} \tag{5.36}$$

and

$$\sigma = \lambda(\nabla \cdot u)I + \mu(\nabla u + (\nabla u)^T). \tag{5.37}$$

Here $u : \Gamma \to \mathbb{R}^3$ is the unknown displacement vector field, subscript t denotes a derivative with respect to time, $\sigma : \Gamma \to \mathbb{R}^{3\times 3}$ is the stress tensor, f is the body force per unit volume, λ and μ are Lamé's elasticity parameters for the material in Γ, I is the

(a) (b)

Figure 5.1 – (a) initial condition and a snapshot of the 3D beam. (b) initial condition and a snapshot of the 2D beam with a cavity.

identity tensor, n is the outward unit normal vector at the boundary and $\tau : \partial\Gamma_\tau \to \mathbb{R}^3$ is the traction at a subset of the boundary $\partial\Gamma_\tau$ [67]. We refer to Figure 5.1.(a) for a snapshot of the elastic beam.

We define a vector valued function space as $V = \{u \in (L^2(\Gamma))^3 : \|\nabla u_i\|_2 \in L^2, i = 1,2,3, u = \vec{0} \text{ on } \partial\Gamma_\tau\}$, equipped with the standard L^2 inner product $(\cdot,\cdot) : V \times V \to \mathbb{R}$, and seek the solution to (5.36). To derive the weak formulation of (5.36), we multiply it with the vector valued test function $v \in V$, integrate over Γ, and use integration by parts to obtain

$$\int_\Gamma u_{tt} \cdot v \, dx = -\int_\Gamma \sigma : \nabla v \, dx + \int_{\partial\Gamma_\tau} (\sigma \cdot n) \cdot v \, ds + \int_\Gamma f \cdot v \, dx, \tag{5.38}$$

where $\sigma : \nabla v = \sum_{i,j} \sigma_{ij}(\nabla v)_{ji}$ is the tensor inner product. Note that the skew-symmetric part of ∇v vanishes over the product $\sigma : \nabla v$, since σ is symmetric. By prescribing the boundary conditions to (5.38) we recover

$$\int_\Gamma u_{tt} \cdot v \, dx = -\int_\Gamma \sigma : \mathrm{Sym}(\nabla v) \, dx + \int_{\partial\Gamma_\tau} \tau \cdot v \, ds + \int_\Gamma f \cdot v \, dx, \tag{5.39}$$

with $\mathrm{Sym}(\nabla v) = (\nabla v + (\nabla v)^T)/2$. The variational form associated to (5.36) is

$$(u_{tt}, v) = -a(u,v) + b(v), \quad u, v \in V, \tag{5.40}$$

where

$$a(u,v) = \int_\Gamma \sigma : \mathrm{Sym}(\nabla v) \, dx, \; b(v) \quad = \int_{\partial\Gamma_\tau} \tau \cdot v \, ds + \int_\Gamma f \cdot v \, dx. \tag{5.41}$$

To obtain the FEM discretization of (5.40), we triangulate the domain Γ and define vector valued piece-wise linear basis functions $\{\phi_i\}_{i=1}^{N_h}$. We define the FEM space V_h, an approximation of V, as the span of those basis functions. Projecting (5.40) onto V_h

yields the discrete weak form

$$((u_h)_{tt}, v_h) = -a(u_h, v_h) + b(v_h), \quad u_h, v_h \in V_h. \tag{5.42}$$

Any particular function u_h can be expressed as $u_h = \sum_{i=1}^{N_h} q_i \phi_i$, where q_i, $i = 1, \ldots, N_h$, are the expansion coefficients. Therefore, by choosing test functions $v_h = \phi_i$, $i = 1, \ldots, N_h$, we obtain the system of ODEs

$$M\ddot{q} = -Kq + g_q. \tag{5.43}$$

where $q = (q_1, \ldots, q_{N_h})^T$ are unknowns, the *mass matrix* $M \in \mathbb{R}^{N_h \times N_h}$ is given as $M_{i,j} = (\phi_i, \phi_j)$, the *stiffness matrix* $K \in \mathbb{R}^{N_h \times N_h}$ is given as $K_{i,j} = a(\phi_j, \phi_i)$ and $g_q = (b(v_1), \ldots, b(v_{N_h}))^T$. Now introduce the canonical coordinate $p = M\dot{q}$ to recover the Hamiltonian system

$$\dot{z} = \mathbb{J}_{2N_h} L z + g_{qp}, \tag{5.44}$$

where

$$z = \begin{pmatrix} q \\ p \end{pmatrix}, \quad L = \begin{pmatrix} K & 0 \\ 0 & M^{-1} \end{pmatrix}, \quad g_{qp} = \begin{pmatrix} 0 \\ g_q \end{pmatrix}, \tag{5.45}$$

together with the Hamiltonian function $H(z) = \frac{1}{2} z^T L z + z^T \mathbb{J}_{2N_h}^T g_{qp}$. An appropriate FEM setup leads to a symmetric and positive-definite matrix L. Hence, it seems natural to take $X = L$, the energy matrix associated to (5.44). The system parameters are summarized in the table below. For further information regarding the problem, we refer to [67].

Domain shape	box: $l_x = 1,\ l_y = 0.2,\ l_z = 0.2$
Time step-size	$\Delta t = 0.01$
Gravitational force	$f = (0, 0, -0.4)^T$
Traction	$\tau = \vec{0}$
Lamé parameters	$\lambda = 1.25$, $\mu = 1.0$
Degrees of freedom	$2N_h = 1650$

Projection operators $P_{X,V}$, $P_{I,A}^{\text{symp}}$ and $P_{X,\tilde{A}}^{\text{symp}}$ are constructed following Algorithms 3.2, 4.1 and 5.1, respectively, with $\delta = 5 \times 10^{-4}, 2 \times 10^{-4}$ and 1×10^{-4}. The reduced systems, obtained from $P_{I,A}^{\text{symp}}$ and $P_{X,\tilde{A}}^{\text{symp}}$, are integrated in time using the Störmer-Verlet scheme to generate the temporal snapshots. The reduced system obtained from $P_{X,V}$ is integrated using a second order implicit Runge-Kutta method. Note that the Störmer-Verlet scheme is not used since the canonical form of a Hamiltonian system

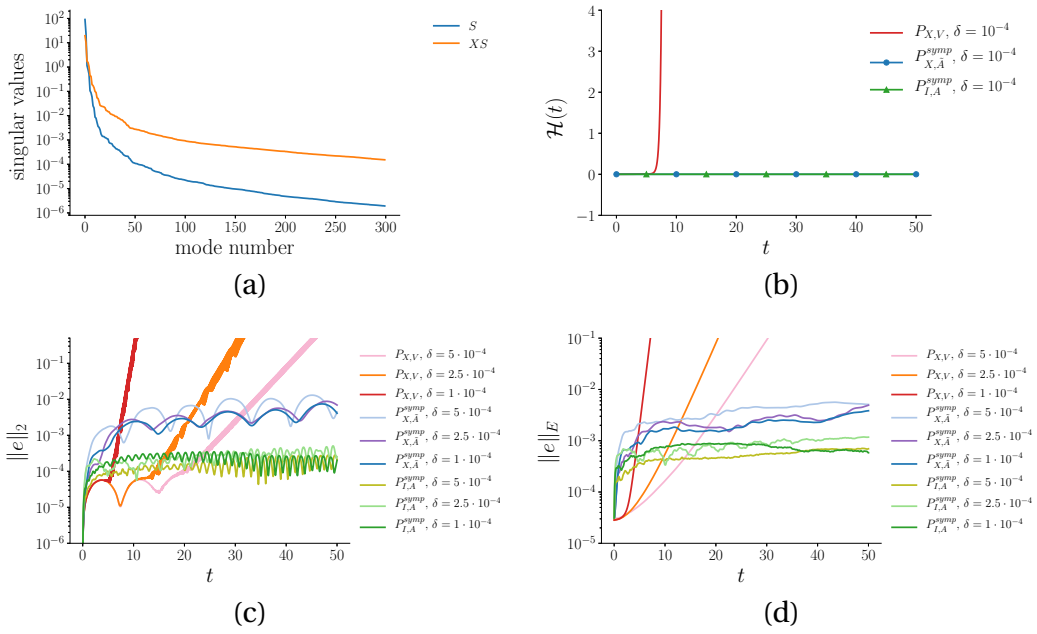


Figure 5.2 – Numerical results related to the beam equation. (a) the decay of the singular values, (b) conservation of the Hamiltonian, (c) error with respect to the 2-norm, (d) error with respect to the X-norm.

is destroyed when $P_{X,V}$ is applied.

Figure 5.2.(a) shows the decay of the singular values of the temporal snapshots S and XS, respectively. The difference in the decay indicates that the reduced systems constructed using $P^{\text{symp}}_{I,A}$ and $P^{\text{symp}}_{X,\tilde{A}}$ would have different sizes for a similar prescribed accuracy.

Figure 5.2.(b) shows the conservation of the Hamiltonian for the methods discussed previously. This confirms that the symplectic methods preserve the Hamiltonian and the system energy. However, the Hamiltonian blows up for the reduced system constructed by the projection $P_{X,V}$.

Figure 5.2.(c) shows the L^2 error between the projected systems and the full order system, defined as

$$\|e\|_{L^2} = \sqrt{(e,e)} \approx \sqrt{(q-\hat{q})^T M (q-\hat{q})}, \tag{5.46}$$

where $e \in V$ is the error function and $\hat{q} \in \mathbb{R}^{2n}$ is an approximation for q. We notice that the reduced system obtained by the non-symplectic method is unstable and the reduced system, constructed using $P_{X,V}$, is more unstable as k increases. On the other hand, the symplectic methods yield a stable reduced system. Although the system, constructed by the projection $P^{\text{symp}}_{X,\tilde{A}}$, is not based on the 2-norm projection, the error

remains bounded with respect to the 2-norm.

We define the energy norm $\|\cdot\|_E : V \to \mathbb{R}$ as

$$\|(u,\dot{u})\|_E = \sqrt{a(u,u)+(\dot{u},\dot{u})} \approx \|z\|_X. \tag{5.47}$$

Figure 5.2.(d) shows the MOR error with respect to the energy norm. We observe that the classical model reduction method based on the projection $P_{X,V}$ does not yield a stable reduced system. However, the symplectic methods provide a stable reduced system. We observe that the original symplectic approach also provides an accurate solution with respect to the energy norm. Nevertheless, the relation between the two norms depends on the problem set up and the choice of discretization [30].

5.6.2 Elastic Beam With Cavity

In this section we investigate the performance of the proposed method on a two dimensional elastic beam that contains a cavity. In this case a nonuniform triangulated mesh is desirable to balance the computational cost of a FEM discretization with the numerical error around the cavity. Figure 5.1.(a) shows the nonuniform mesh used in this section. System parameters are taken to be identical to those in Section 5.6.1. Numerical parameters are summarized in the table below.

cavity width	$l_c = 0.1$
Time step-size	$\Delta t = 4 \times 10^{-4}$
Degrees of freedom	$2N_h = 744$

Figure 5.3.(a) shows the decay of the singular values for the snapshot matrix S and XS. The divergence of the two curves indicates that to obtain the same accuracy in the reduced system, the basis constructed from S and XS would have different sizes. Projection operators $P_{X,A}$, $P^{\text{symp}}_{I,A}$ and $P^{\text{symp}}_{X,\tilde{A}}$ are constructed according to the Algorithms 3.2, 4.1 and 5.1. The error tolerated is set to $\delta = 2.5 \times 10^{-3}$, $\delta = 1 \times 10^{-3}$ and $\delta = 5 \times 10^{-4}$.

The 2-norm error and the error in the energy norm are presented in Figure 5.3.(c) and Figure 5.3.(d), respectively. We notice that although the non-symplectic method is bounded, it results in larger errors compared to the symplectic methods. Moreover, we notice that the error generated by the symplectic methods is consistently reduced under basis enrichment. It is observed that in the energy norm, the projection $P^{\text{symp}}_{X,\tilde{A}}$ provides a more accurate solution by comparing to Figure 5.2. This is due to the nonuniform mesh on which the weight matrix X associates higher weights to the

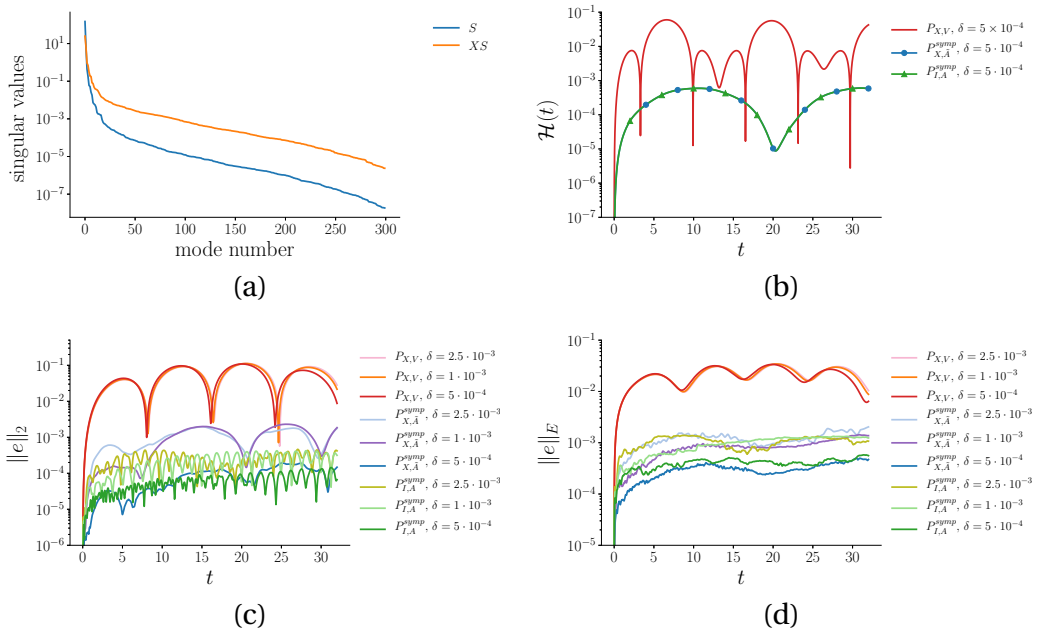


Figure 5.3 – Numerical results related to the beam with cavity. (a) the decay of the singular values, (b) conservation of the Hamiltonian, (c) error with respect to the 2-norm, (d) error with respect to the energy norm.

elements that are subject to larger error. Therefore, we expect the reduced system constructed with the projection $P^{\text{symp}}_{X,\tilde{A}}$, to outperform the one constructed with $P^{\text{symp}}_{I,A}$ on a highly nonuniform mesh.

Figure 5.3.(b) shows the error in the Hamiltonian. Comparing to Figure 5.2, we notice that the energy norm strengthens the boundedness of the non-symplectic method. However, the symplectic methods preserves the Hamiltonian at a higher accuracy

5.6.3 The sine-Gordon equation

The sine-Gordon equation arises in differential geometry and quantum physics [73], as a nonlinear generalization of the linear wave equation of the form

$$\begin{cases} u_t(t,x) = v, \quad x \in \Gamma, \\ v_t(t,x) = u_{xx} - \sin(u), \\ u(t,0) = 0, \\ u(t,l) = 2\pi. \end{cases} \tag{5.48}$$

Here $\Gamma = [0, l]$ is a line segment and $u, v : \Gamma \to \mathbb{R}$ are scalar functions. The Hamiltonian associated with (5.48) is

$$H(q,p) = \int_\Gamma \frac{1}{2}v^2 + \frac{1}{2}u_x^2 + 1 - \cos(u)\, dx. \tag{5.49}$$

One can verify that $u_t = \delta_v H$ and $v_t = -\delta_u H$, where δ_v, δ_u are standard variational derivatives. The sine-Gordon equation admits the soliton solution [73]

$$u(t,x) = 4\text{arctan}\left(\exp\left(\pm\frac{x - x_0 - ct}{\sqrt{1-c^2}}\right)\right), \tag{5.50}$$

where $x_0 \in \Gamma$ and the plus and minus signs correspond to the *kink* and the *anti-kink* solutions, respectively. Here c, $|c| < 1$, is the wave speed. We discretize the segment into n equi-distant grid point $x_i = i\Delta x$, $i = 1, \ldots, n$. Furthermore, we use a standard finite-differences scheme to discretize (5.48) and obtain

$$\dot{z} = \mathbb{J}_{2n} L z + \mathbb{J}_{2n} g(z) + \mathbb{J}_{2n} c_b. \tag{5.51}$$

Here $z = (q^T, p^T)^T$, $q(t) = (u(t,x_1), \ldots, u(t,x_N))^T$, $p(t) = (v(t,x_1), \ldots, v(t,x_N))^T$, c_b is the term corresponding to the boundary conditions and

$$L = \begin{pmatrix} D_x^T D_x & 0_N \\ 0_N & I_n \end{pmatrix}, \quad g(z) = \begin{pmatrix} \sin(q) \\ \vec{0} \end{pmatrix}, \tag{5.52}$$

where D_x is the standard matrix differentiation operator. We may take $X = L$ as the weight matrix associated to (5.51). The discrete Hamiltonian takes the form

$$H_{\Delta x} = \Delta x \cdot \frac{1}{2}\|p\|_2^2 + \Delta x \cdot \|D_x q\|_2^2 + \sum_{i=1}^{n} \Delta x \cdot (1 - \cos(q_i)). \tag{5.53}$$

The system parameters are given as

Domain length	$l = 50$
No. grid points	$n = 500$
Time step-size	$\Delta t = 0.01$
Wave speed	$c = 0.2$

The midpoint scheme (2.38) is used to integrate (5.48) in time and generate the snapshot matrix S. Similar to the previous subsection, projection operators $P_{X,V}$, $P_{I,A}^{\text{symp}}$ and $P_{X,\tilde{A}}^{\text{symp}}$ are used to construct a reduced system. To accelerate the evaluation of the nonlinear term, the symplectic DEIM and the generalized symplectic DEIM, Algorithm 5.2, are coupled with the projection operators $P_{I,A}^{\text{symp}}$ and $P_{X,A}^{\text{symp}}$, respectively.

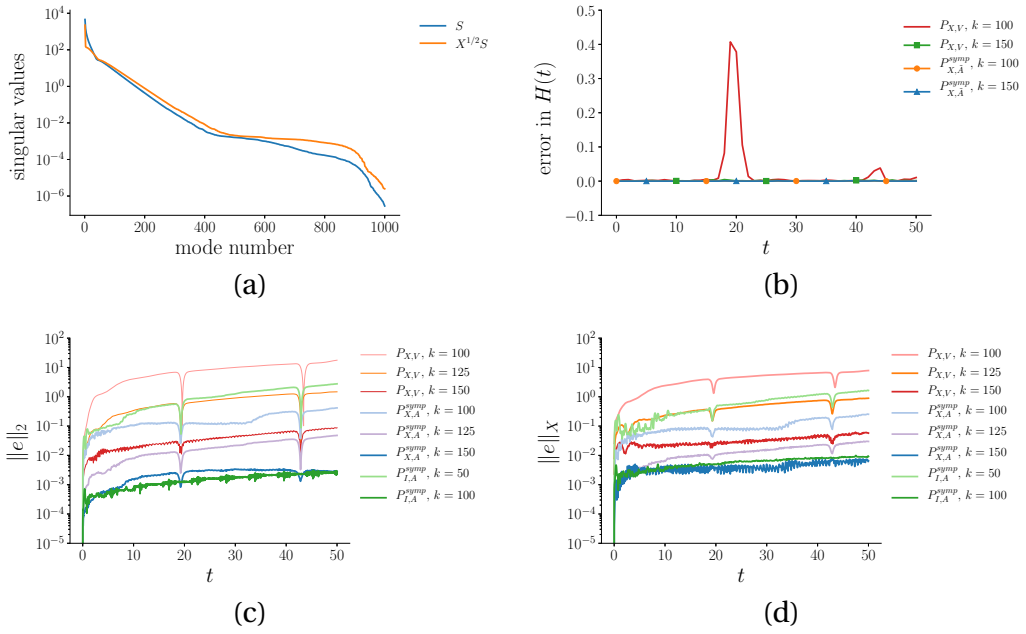


Figure 5.4 – Numerical results related to the sine-Gordon equation. (a) the decay of the singular values, (b) the error in the Hamiltonian, (c) the error with respect to the 2-norm, (d) the error with respect to the energy norm.

Furthermore, the DEIM approximation is used for the efficient evaluation of the reduced system, obtained by the projection $P_{X,V}$. The midpoint rule is also used to integrate the reduced systems in time. Figure 5.4 shows the numerical results obtained with the reduced models without approximating the nonlinearity, while the results for the accelerated evaluation of the nonlinear term are presented in Figure 5.5.

Figure 5.4.(a) shows the decay of the singular values of matrices S and XS. As previously, we observe a saturation in the decay of the singular values of XS compared to the singular values of S. This indicates that the reduced basis, based on a weighted inner product, should be chosen to be larger to provide an accuracy similar to based on the Euclidean inner product.

Put differently, unweighted reduced bases, when compared to the weighted ones, may be highly inaccurate in reproducing the underlying physical properties of the system.

Figure 5.4.(b) displays the error in the Hamiltonian. It is again observed that the symplectic approaches conserve the Hamiltonian. However, the classic approaches do not necessarily conserve the Hamiltonian. We point out that using the projection operator $P_{X,V}$ ensures the boundedness of the Hamiltonian. The contrary is observed when we apply the POD with respect to the Euclidean inner-product, i.e. applying the projection operator $P_{I,V}$. This can be seen in the results presented in [81], where the unboundedness of the Hamiltonian is observed when $P_{I,V}$ is applied to the sine-

Gordon equation. Nevertheless, only the symplectic model reduction consistently preserves the Hamiltonian.

Figure 5.4.(c) shows the error with respect to the Euclidean inner-product between the solution of the projected systems and the original system. The behavior of the solution is investigated for $k = 100$, $k = 125$ and $k = 150$. We observe that all systems that are projected with respect to the X-norm are bounded. As the results in [81] suggest, the Euclidean inner-product does not necessarily yield a bounded reduced system. Moreover, we notice that the symplectic projection $P^{\text{symp}}_{X,\tilde{A}}$ results in a substantially more accurate reduced system compared to the reduced system yielded from $P_{X,V}$. This is because the overall behavior of the original system is translated correctly to the reduced system through the symplectic projection.

The error with respect to the X-norm between the solution of the original system and the projected systems is presented in Figure 5.4.(d). We observe that the behavior of the X-norm error is similar to that in the Euclidean norm. However, the growth of the error is slower for methods based on a weighted inner product. Note that the connection between the error in the Euclidean norm and the X-norm is problem and discretization dependent. We also observe that symplectic methods are substantially more accurate.

Figure 5.5 shows the performance of the different model reduction methods, when an efficient method is adopted for evaluating the nonlinear term in (5.51). This figure compares the symplectic approaches against non-symplectic methods. For all simulations, the size of the reduced basis for (5.51) is chosen to be $k = 100$. The size of the basis of the nonlinear term is taken as $k_n = 75$ and $k_n = 100$. For symplectic methods, a basis for the nonlinear term is constructed according to Algorithm 5.2, whereas for non-symplectic methods, the DEIM is applied. Note that for symplectic methods, the basis for the nonlinear term is added to the symplectic basis A. This means that the size of the reduced system is larger when compared to the classical approach.

5.7 Conclusion

This chapter presents a model reduction approach that combines the classic model reduction method, defined with respect to a weighted inner product, with symplectic model reduction. This allows the reduced system to be defined with respect to norms and inner-products that are natural to the problem and most suitable for the method of discretization. Furthermore, the symplectic nature of the reduced system preserves the Hamiltonian structure of the original system, which results in robustness and enhanced stability in the reduced system.

It is demonstrated that including the weighted inner-product in the symplectic model

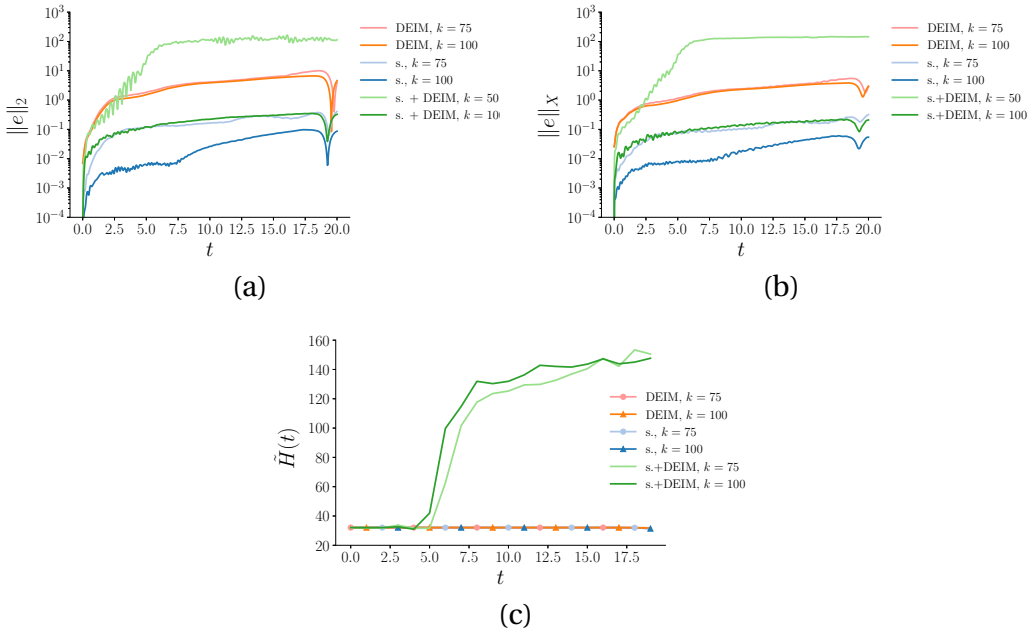


Figure 5.5 – Numerical results related to the sine-Gordon equation with efficient evaluation of the nonlinear terms. Here, "DEIM" indicates classical model reduction with the DEIM, "s.+DEIM" indicates symplectic model reduction with the DEIM and "s." indicates symplectic model reduction with symplectic treatment of the nonlinear term. (a) error with respect to the Euclidean norm, (b) error with respect to the X-norm, (c) error in the Hamiltonian.

reduction can be viewed as a natural extension of the unweighted symplectic method. Therefore, the stability preserving properties of the symplectic method generalize naturally to the new method.

Numerical results suggest that classic model reduction methods with respect to a weighted inner product can help with the boundedness of the system. However, only the symplectic treatment can consistently increase the accuracy of the reduced system. This is consistent with the fact the symplectic methods preserve the Hamiltonian structure.

We also show that to accelerate the evaluation of the nonlinear terms, adopting a symplectic approach is essential. This allows an accurate reduced model that is consistently improving when the basis for the nonlinear term is enriched.

Hence, the symplectic model-reduction with respect to a weighted inner product provides an accurate and robust reduced system that allows the use of the norms and inner products most appropriate to the problem.

6 Symplectic Model Order Reduction of Dissipative Hamiltonian Systems

It was discussed in Chapters 4 and 5, that if the basis for the reduced space is not chosen carefully, the symplectic symmetry of Hamiltonian will be destroyed by model reduction. To resolve this issue, the symplectic MOR constructs a reduced order configuration space that inherits symmetries of the full configuration space. By using a proper time integrator scheme, the symmetries are preserved in the reduced system. A greedy-type algorithm is developed in Algorithms 4.1 and 5.1 for construction of a basis for such a reduced configuration space.

Most models in engineering appear as a dissipative perturbation of a Hamiltonian system. In these systems, conservation of energy is taken as a fundamental principle of the system dynamics, while dissipative forces, e.g. friction, can change the energy of the system [106]. As the energy is no longer preserved for such systems, existing methods can no longer be applied directly [81].

For dissipative and forced Hamiltonian system, Peng et al. [80] suggest a symplectic model reduction method that preserves the Hamiltonian and the dissipative structure of the original system. However, since this method uses a symplectic integrator for a non-conservative system, there is no guarantee that the evolution of the energy is translated correctly to the reduced system.

In the context of network modeling and circuit simulation, considerable work has been done in the development of structure preserving, and in particular energy preserving, model reduction techniques. Model reduction for port-Hamiltonian systems are given in [84, 9, 26] and the references therein. These methods use a Krylov or a Proper Orthogonal Decomposition (POD) approach to construct a reduced port-Hamiltonian system that preserves the passivity, and, thus, the stability of the original system. However, these methods do not generally guarantee the correct distribution of the energy among the energy consuming and energy storing units. Furthermore, over long time integration, accumulation of local errors might produce an erroneous solution.

This chapter presents the reduced dissipative Hamiltonian (RDH) method as a structure preserving model reduction approach for dissipative Hamiltonian systems. A key difference between this method and the other existing methods is that the RDH enables the reduced system to be integrated using a symplectic integrator. By considering a canonical heat bath, also known as hidden strings [40, 39], the reduced system is extended to a closed and conservative system. Therefore, a symplectic time integrator can be used to guarantee conservation of the system energy and the correct dissipation of energy. Furthermore, the hidden strings assure that the local errors in the dissipation of energy do not accumulate, resulting in a correct evolution of the system energy.

This chapter is organized as follows: Section 6.1 covers the required background on dissipative Hamiltonian systems and the Hamiltonian extension. In Section 6.2 we introduce the reduced dissipative Hamiltonian (RDH) method. Accuracy, stability and efficiency of the RDH method is discussed in Section 6.3, and illustrated through simulation of the dissipative wave equation and a linear port-Hamiltonian system of an electrical circuit. We offer conclusive remarks in Section 6.4.

6.1 Dissipative Hamiltonian Systems and Hamiltonian Extensions

Many systems in engineering and science appear as a perturbation of a Hamiltonian system, where the perturbation can be regarded as dissipation. In these systems, the energy tends to decrease over time, and thus, the symmetry expressed in Proposition 2.10 and the conservation law in Corollary 2.11 are violated. Therefore, it is common to take the conservation of energy as a fundamental principle and consider the dissipative system coupled with a heat bath that absorbs the dissipated energy of the original system.

To account for dissipation in a quadratic Hamiltonian $H(z) = \frac{1}{2} z^T K^T K z$, we reformulate a Hamiltonian system as a time dispersive and dissipative (TDD) [40] system

$$\begin{cases} \dot{z} = \mathbb{J}_{2n} K^T f(t, z), \\ z(0) = z_0, \end{cases} \tag{6.1}$$

where f is the solution to the Volterra integral equation [58, 28]

$$f(t, z) + \int_0^t \chi(t-s) f(s, z)\, ds = Kz. \tag{6.2}$$

Here $\chi : \mathbb{R}^+ \to \mathbb{R}^{2n \times 2n}$ is a bounded matrix valued function with respect to the Frobenius norm and is called the *general susceptibility*. Note that the integral term in (6.2)

accounts to the accumulation of the dissipation, whereas $\chi(s) = 0$ implies (6.1) is equivalent to (4.1). Furthermore, under suitable assumptions on K, both (4.1) and (6.1) are well-posed [40].

Example 6.1. *Consider the dynamics of the damped harmonic oscillator*

$$\ddot{q} + r\dot{q} + kq = 0 \tag{6.3}$$

where k is the Hooke's constant and r is the spring's damping factor. Note that without a damping term, (6.3) is a Hamiltonian system. The TDD formulation for the damped harmonic oscillator takes the form (6.1) with

$$z = \begin{pmatrix} q \\ \dot{q} \end{pmatrix}, \quad K = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{k} \end{pmatrix}, \quad \chi(t) = r. \tag{6.4}$$

Here $\mathbb{J}_{2n} = \mathbb{J}_2$, $(q, \dot{q})^T$ is the canonical coordinate and the susceptibility is the constant function r.

It is shown in [40, 39] that under natural assumptions on the linear susceptibility $\chi(t)$ (see below), one can couple a TDD system of the form (6.1) with a canonical heat bath where the dissipated energy is captured in the heat bath in a canonical sense. In other words, one can construct a Hilbert space $\mathcal{H}$ and an isometric injection $I : \mathbb{R}^{2n} \to \mathbb{R}^{2n} \times \mathcal{H}^{2n}$ where the solution z to (6.1) is the projection of x onto $\mathbb{R}^{2n}$, and x is the solution to

$$\dot{x} = \mathcal{J}_{2n} \frac{\delta H_{\text{ex}}}{\delta x}. \tag{6.5}$$

Here, $x \in \mathbb{R}^{2n} \times \mathcal{H}^{2n}$, $H_{\text{ex}} : \mathbb{R}^{2n} \times \mathcal{H}^{2n} \to \mathbb{R}$ is an extended quadratic Hamiltonian function and $\mathcal{J}_{2n}$ is the symplectic operator defined on $\mathbb{R}^{2n} \times \mathcal{H}^{2n}$ respectively.

Theorem 6.1. *[40] Suppose that K is full rank and $\chi(t)$ is symmetric. Then there is a quadratic extension to (6.1) of the form (6.5), if*

$$Im(\xi\hat{\chi}(\xi)) \geq 0, \quad \forall \xi = \omega + i\eta,\ \eta \geq 0, \tag{6.6}$$

where $\hat{\chi}$ is the Fourier-Laplace transform of χ

$$\hat{\chi}(\xi) = \int_0^\infty e^{i\xi t}\chi(t)\, dt. \tag{6.7}$$

When χ is a constant symmetric matrix, condition (6.6) corresponds to χ being positive

semi-definite [40]. In this case the Hamiltonian extension to (6.1) reads

$$\dot{z}(t) = \mathbb{J}_{2n} K^T f(t), \tag{6.8a}$$
$$\partial_t \phi(t, x, , z) = \theta(t, x, z), \tag{6.8b}$$
$$\partial_t \theta(t, x, z) = \partial_x^2 \phi(t, x, z) + \sqrt{2}\delta_0(x) \cdot \sqrt{\chi} f(t, z), \tag{6.8c}$$

together with the initial condition

$$z(0) = z_0, \quad \phi(0, \cdot, \cdot) = 0, \quad \theta(0, \cdot, \cdot) = 0. \tag{6.9}$$

Here θ and ϕ are vector valued functions in $\mathcal{H}^{2n}$, $\delta_0(s)$ is the Dirac's delta function, $\sqrt{\chi}$ is the matrix square root of χ and f is the solution to the equation

$$f(t, z) + \sqrt{2} \cdot \sqrt{\chi} \phi(t, 0, z) = Kz(t). \tag{6.10}$$

Equations (6.8b) and (6.8c) are equations for a vibrating string, and can be solved analytically

$$\phi(t, x, z) = \frac{\sqrt{2}}{2} \int_0^{t-|x|} \sqrt{\chi} f(s, z)\, ds, \quad \theta(t, x, z) = \frac{\sqrt{2}}{2} \cdot \sqrt{\chi} f(t - |x|, z). \tag{6.11}$$

We can recover (6.2) by substituting (6.11) into (6.10). The extended Hamiltonian H_{ex} for the system Equations (6.8a) to (6.8c) takes the quadratic form

$$H_{\text{ex}}(z, \phi, \theta) = \int_\Gamma \frac{1}{2} \left(\|Kz - \phi(t, 0, z)\|_2^2 + \|\theta(t, z)\|_{\mathcal{H}^{2n}}^2 + \|\partial_x \phi(t, z)\|_{\mathcal{H}^{2n}}^2 \right) dz, \tag{6.12}$$

where $z \in \Gamma$, $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^{2n}$ and $\|\cdot\|_{\mathcal{H}^{2n}}$ is the induced norm from the inner product on $\mathcal{H}^{2n}$.

Equations (6.8b) and (6.8c) are called the *hidden strings*. The dissipation of energy in the original system (6.1) is carried away, as vibrations, along the added strings making the extended system conservative. The Hamiltonian extension of the damped harmonic oscillator in Example 6.1 is exactly the Lamb model [66] which is a harmonic oscillator coupled with a vibrating string, and the tension in the string causes linear dissipation in the dynamics of the harmonic oscillator.

Note that the time integration of Equations (6.8a) to (6.8c) involves the integration of f in (6.11). In general, the history of $f(t, z)$ must be stored and may cause storage limitation in long-time integration. However, we are interested solely in finding $z(t, z)$ which depends on f at time t, and $\phi(t, 0, z)$, i.e. the integral of the history of f. So by carefully choosing a quadrature rule that uses the same quadrature nodes as the time integrator we can avoid storing the history of f. For example for the trapezoidal rule,

we recover the recursive relation

$$\int_0^{t_n} f(s,z)\,ds \approx \frac{\Delta t}{2} f(t_n,z) + \frac{\Delta t}{2} f(t_{n-1},z) + \int_0^{t_{n-1}} f(s,z)\,ds, \tag{6.13}$$

where Δt is the time step. The recursive relation in (6.13) suggests that storing the value of the integral term together with the state of f in the previous time step suffices to evaluate the integral for the new time step. For other interpolation based quadrature rules, we can construct similar recursive rules of the form

$$\int_0^{t_n} f(s,z)\,ds \approx \sum_{i=0}^{k} \omega_i f(t_{n-i},z) + \int_0^{t_{n-k}} f(s,z)\,ds \tag{6.14}$$

for some quadrature weights ω_i, $i = 1,\ldots,k$ with $k \ll n$. Thus, time integration of Equations (6.8a) to (6.8c), only requires storage of k evaluations of f.

6.2 The Reduced Dissipative Hamiltonian Method

Since the symplectic model reduction discussed in Chapters 4 and 5 are based on the conservation law in Theorem 2.11, it can no longer be applied to dissipative Hamiltonian systems. Instead in the reduced dissipative Hamiltonian method, we consider a Hamiltonian extension to a dissipative Hamiltonian system to construct a closed system. A symplectic model reduction can then be naturally applied to conserve the total energy.

Consider a dissipative Hamiltonian system of the form (6.1) with a quadratic Hamiltonian, $H(z) = z^T K^T K z$. Since $K^T K$ is symmetric and positive definite, it has a unique Cholesky factorization $K^T K = L^T L$ where L is upper triangular [100]. So we can write

$$H(z) = z^T L^T L z. \tag{6.15}$$

Further, suppose that the solution $z(t)$ lies on a low-dimensional symplectic subspace such that $z \approx Ay$, where A is an ortho-symplectic matrix of the size $2n \times 2k$ satisfying condition (d) in Proposition 4.1, and y is the expansion coefficients of z in the basis of A. Writing (6.1) in terms of the reduced coordinates y reads

$$A\dot{y}(t) = \mathbb{J}_{2n} L^T f(t, Ay) + r(z), \tag{6.16}$$

together with the complementary equation

$$f(t,Ay) + \sqrt{2}\cdot\sqrt{\chi}\phi(t,0,y) = LAy. \tag{6.17}$$

The symplectic Galerkin projection implies

$$\dot{y}(t) = \mathbb{J}_{2k} A^T L^T f(t, Ay), \tag{6.18}$$
$$A^T f(t, Ay) + \sqrt{2} A^T \sqrt{\chi} \phi(t, 0, Ay) = A^T L A y, \tag{6.19}$$

where we use the fact that $A^+ \mathbb{J}_{2n} = \mathbb{J}_{2k} A^T$. If we define

$$f = A\tilde{f}, \quad \phi = A\tilde{\phi}, \quad \theta = A\tilde{\theta}, \quad \tilde{L} = A^T L A,$$

and the *reduced susceptibility* as $\tilde{\chi} = A^T \chi A$ we recover the reduced Hamiltonian system

$$\dot{y}(t) = \mathbb{J}_{2k} \tilde{L}^T \tilde{f}(t, y), \tag{6.20a}$$
$$\partial_t \tilde{\phi}(t, x, y) = \tilde{\theta}(t, x, y), \tag{6.20b}$$
$$\partial_t \tilde{\theta}(t, x, y) = \partial_x^2 \tilde{\phi}(t, x, y) + \sqrt{2} \delta_0(x, y) \cdot \sqrt{\tilde{\chi}} \tilde{f}(t, y), \tag{6.20c}$$

together with the auxiliary equation

$$\tilde{f}(t, y) + \sqrt{2} \sqrt{\tilde{\chi}} \tilde{\phi}(t, 0, y) = \tilde{L} y. \tag{6.21}$$

Equations (6.20a) to (6.20c) is a Hamiltonian system on the symplectic linear vector space $\mathbb{R}^{2k} \times \mathcal{H}^{2k}$ and contributes to the *reduced TDD system*

$$\dot{y} = \mathbb{J}_{2k} \tilde{L}^T \tilde{f}(t, y), \quad \tilde{f}(t, y) + \int_0^t \tilde{\chi} \cdot \tilde{f}(s, y)\, ds = \tilde{L} y. \tag{6.22}$$

with the Hamiltonian defined as

$$\tilde{H}(y) = y^T \tilde{L}^T \tilde{L} y. \tag{6.23}$$

Therefore, the system energy will be conserved along integral curves of Equations (6.20a) to (6.20c).

We point out that the transformation that connects Equations (6.8a) to (6.8c)) to Equations (6.20a) to (6.20c) is given by

$$\mathbf{A} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} : \mathbb{R}^{2n} \times \mathcal{H}^{2n} \to \mathbb{R}^{2k} \times \mathcal{H}^{2k}. \tag{6.24}$$

This is a symplectic transformation, since $\mathbf{A}^T \mathcal{J}_{2n} \mathbf{A} = \mathcal{J}_{2k}$. Furthermore, the dissipation of energy in the reduced system only depends on the reduced susceptibility. Thus, the choice of A should be independent of the hidden strings (ϕ, θ). In other words, if the reduced space is chosen to be a symplectic subspace, then the actions of model reduction and Hamiltonian extension commute. We summarize the algorithm for

model reduction of dissipative Hamiltonian systems in Algorithm 6.1.

Algorithm 6.1 The Reduced Dissipative Hamiltonian Method (RDH)

1: Construct the Hamiltonian extension Equations (6.8a) to (6.8c)) to the original TDD system (6.1).
2: Collect the snapshots $z(t_i)$, $i = 1, \ldots, N$ through time integration of the extended Hamiltonian.
3: Construct an ortho-symplectic basis A.
4: Define $\tilde{L} = A^T L A$, $\tilde{\chi} = A^T \chi A$ and construct the reduced dissipative Hamiltonian system Equations (6.20a) to (6.20c)

Note that Algorithm 6.1 does not depend on the choice of the method to construct an ortho-symplectic basis A. Thus, for basis generation, the symplectic greedy or any SVD-based methods discussed in Chapter 4 can be used.

The main advantage of the RDH method compared to the existing methods is that it enables the reduced system to be integrated using a symplectic integrator. The reduced system constructed using the RDH is a closed Hamiltonian system, therefore the conservation law in Corollary 2.11 holds and a symplectic integrator guarantees that the total energy is preserved in the reduced system. Alternative methods, e.g. [80, 84, 9], either integrate the reduced system with a non-symplectic integrator, or do not construct a closed reduced system which result in accumulation of local errors or unstable solution during long time integration, respectively [51].

The following theorem provides an indication for the boundedness of the reduced system.

Theorem 6.2. *Consider the TDD system (6.1) together with the initial condition (6.9). If in the absence of dissipation* ($\chi = 0$)*, a bounded open neighborhood* $U \subset \mathbb{R}^{2n}$ *of* z_0 *exists such that* $H(z) < c$*, for all* $z \in U$*, and for some constant value* $c > H(z_0)$*, then the RDH method yields a bounded reduced system.*

Proof. For a Hamiltonian of the form (6.15), $H(z) \geq 0$ for all $z \in \mathbb{R}^{2n}$. It is shown in [40, 39] that for a symmetric and positive definite χ, $\frac{d}{dt}H(z(t)) \leq 0$. Therefore, for all $0 \leq t < \infty$, $H(z) < c$ which implies $z(t) \in U$. Now suppose that the symplectic basis A is used to construct the RDH reduced system (6.20a)-(6.20c) such that $y_0 = A^+ z_0$. It is easily checked that $V = \text{Range}(A) \cap U$ is a bounded open neighborhood of z_0 and that $H(z) = H(Ay) < c$ for all $z \in V$. Since $\tilde{H}(y) = H(Ay)$ is a Hamiltonian function for the dissipative Hamiltonian system (6.22) and that $\tilde{\chi}$ is symmetric and positive definite, then $\frac{d}{dt}\tilde{H}(y) \leq 0$. This implies that $\tilde{H}(y) < c$ and $y(t) \in U$ for all $0 \leq t < \infty$. Also it is easy to check that the boundedness of the solution of the original TDD system implies boundedness of the extended system in $\mathbb{R}^{2n} \times \mathcal{H}^{2n}$ with respect to the norm $\|(x, f)\|_{\text{ex}} = \|x\|_2 + \|f\|_{\mathcal{H}^{2n}}$. □

Note that in practice, ensuring $\frac{d}{dt}H(z(t)) \leq 0$, requires a careful time integration of the dissipative system. The symplectic time integration of the extended system guarantees the correct evolution of the Hamiltonian.

We conclude this section by showing that the RDH method preserves the stability of the equilibrium points of the extended system. This provides a strong indication that the overall dynamics and stability of the original system system is translated into the RDH system.

Theorem 6.3. *[2] Consider a Hamiltonian system of the form Equations* (6.8a) *to* (6.8c) *with a Hamiltonian $H \in C^2$ together with the reduced system Equations* (6.20a) *to* (6.20c)*. Suppose that (z_e, ϕ_e, θ_e) is a strict local minimum of H_{ex}. If we can find an open ball neighborhood S of (z_e, ϕ_e, θ_e) with respelct to the norm $\|\cdot\|_{ex}$ such that Range$(\mathbf{A}) \cap S \neq \emptyset$, then the reduced system* (4.10) *remains bounded.*

Proof. Since $H \in C^2$ and H_{ex} is a quadratic extention of H, then H_{ex} is continuously differentiable with respect to all variables (z, ϕ, θ). Furthermore, we can assume that there is a bounded neighborhood S of (z_e, ϕ_e, θ_e) with respect to the norm $\|\cdot\|_{\text{ex}}$ such that $H(z) < c < \infty$ for all $z \in S$, and $\inf_{z\in\partial S} H = c$. Theorem 6.2 suggests that the reduced system remains bouded in Range$(\mathbf{A}) \cap S \neq \emptyset$ □

6.3 Numerical Results

In the following we illustrate the performance of the method through the reduced order model of the dissipative wave equation and a port-Hamiltonian model for a dissipative circuit.

6.3.1 Dissipative wave equation

Consider the dissipative linear wave equation

$$\begin{cases} q_t(t,x) = p(t,x), \\ p_t(t,x) = c^2 q_{xx}(t,x) - r(x)p(t,x), \\ q(0,x) = q_0(x), \\ p(0,x) = 0. \end{cases} \tag{6.25}$$

where x belongs to a one-dimensional torus of length L and $r : [0,1] \to [0,1]$ is a positive semi-definite real valued function.

We discretize the torus into $N_{\Delta x}$ equidistant points and define $\Delta x = L/N_{\Delta x}$, $x_i = i\Delta x$, $q_i = q(t, x_i)$ and $p_i = p(t, x_i)$ for $i = 1, \ldots, N_{\Delta x}$. The discretization of r corresponds to a

diagonal and semi-positive definite matrix r_Δ. Furthermore, we discretize (6.25) using a standard central finite differences schemes to obtain

$$\dot{z} = \mathbb{J}_{2n} K^T K z - R z, \tag{6.26}$$

where $z = (q_1, \ldots, q_{N_{\Delta x}}, p_1, \ldots, p_{N_{\Delta x}})$ and K and R are given as

$$K^T K = \begin{pmatrix} I & 0 \\ 0 & c^2 D_x^T D_x \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 0 \\ 0 & r_\Delta \end{pmatrix}, \tag{6.27}$$

with $D_x^T D_x = D_{xx}$ as the central finite differences matrix operator. Writing (6.26) in a TDD formulation yields

$$\dot{z} = \mathbb{J}_{2n} K^T f(t), \quad f(t, z) + R \int_0^t f(s, z)\, ds = K z. \tag{6.28}$$

Since R is not time dependent, it commutes with the integration operator. The Hamiltonian extension of (6.28), then takes the form (6.8a) to (6.8c).

The initial condition used is given by

$$q_i(0) = h(10 \times |x_i - \frac{1}{2}|), \quad p_i = 0, \quad i = 1, \ldots, N, \tag{6.29}$$

where $h(s)$ is the cubic spline function

$$h(s) = \begin{cases} 1 - \frac{3}{2}s^2 + \frac{3}{4}s^3, & 0 \le s \le 1, \\ \frac{1}{4}(2-s)^3, & 1 < s \le 2, \\ 0, & s > 2. \end{cases} \tag{6.30}$$

For the numerical time integration of the extended Hamiltonian system, the Strömer-Verlet time stepping scheme (2.36) is used. In each time step, the system of linear equations (6.10) is solved to recover z. System parameters are summarized below.

Domain length	$L = 1$
No. grid points	$N = 500$
Space discretization size	$\Delta x = 0.002$
Time discretization size	$\Delta t = 0.002$
Wave speed	$c^2 = 0.1$

The first numerical experiment corresponds to an inhomogeneous dissipative media. Here, r_Δ is a diagonal matrix with diagonal elements $r_i := 0.1 + 0.9(i/N_{\Delta x})$, for $i =$

$1, \ldots, N_{\Delta x}$.

Figure 6.1.(a) shows the solution of the original dissipative wave equation (6.25) at $t \in \{0, 2.5, 5, 7.5\}$. For a nonzero r_Δ the solution will converge to $(q(t = \infty, x), p(t = \infty, x)) = (\rho, 0)$ where ρ is the center of mass of q_0.

We construct the RDH reduced system according to the Algorithm 6.1 using a basis constructed by the greedy basis selection described in Algorithm 4.1. The performance of the method is then compared to the POD and the PSD proposed in [80]. The PSD method constructs a symplectic basis using the cotangent lift method [80]. Note the cotangent lift method can also be used to construct a basis for the RDH method. However, the greedy method and the cotangent lift yield very similar results, inline with the results in previous works [2].

Figure 6.1.(b) illustrates the decay of the singular values of the snapshot matrix [52], for the POD, PSD, and the RDH methods. Note that the snapshots for the PSD and the RDH are different since they have different canonical representations. The fast decay of the eigenvalues in all methods is a strong indicator for the existence of a low dimensional reduced system. The reduced bases are then constructed using 20, 40 and 60 number of modes.

The L^2-error between the full system and the RDH, the PSD, and the POD methods are presented in Figure 6.1.(c). We notice that the symplectic methods provide a more accurate solution when compared to the POD method. In fact, the POD method does not yield a stable reduced system. Furthermore, it is seen that enriching the PSD reduced basis does not yield a significant improvement in the accuracy of the reduced system. This happens as the PSD method, numerically integrate a non-conservative system with a symplectic integrator. This results in an incorrect evolution of the energy and eventually, in a qualitatively wrong numerical solution.

On the other hand, we notice that the RDH method with 40 modes provides a significantly more accurate solution compared to the PSD method with 60 modes. The RDH method provides a conservative reduced system where the dissipated energy is absorbed by the hidden strings and the conservation of the energy is then guaranteed by using a symplectic integrator. Therefore, we observe remarkable increase in the accuracy by enriching the RDH reduced basis.

Figure 6.1.(d) shows the conservation of the energy in the different methods. The conservation law expressed in Theorem 2.1 is destroyed through the POD model reduction and as a consequence we observe blow-up of the system energy. The symplectic methods preserves the energy significantly better. As discussed above, enriching the PSD basis does not significantly improve the preservation of energy. On the contrary, the RDH provides a substantial improvement in the accuracy of the energy.

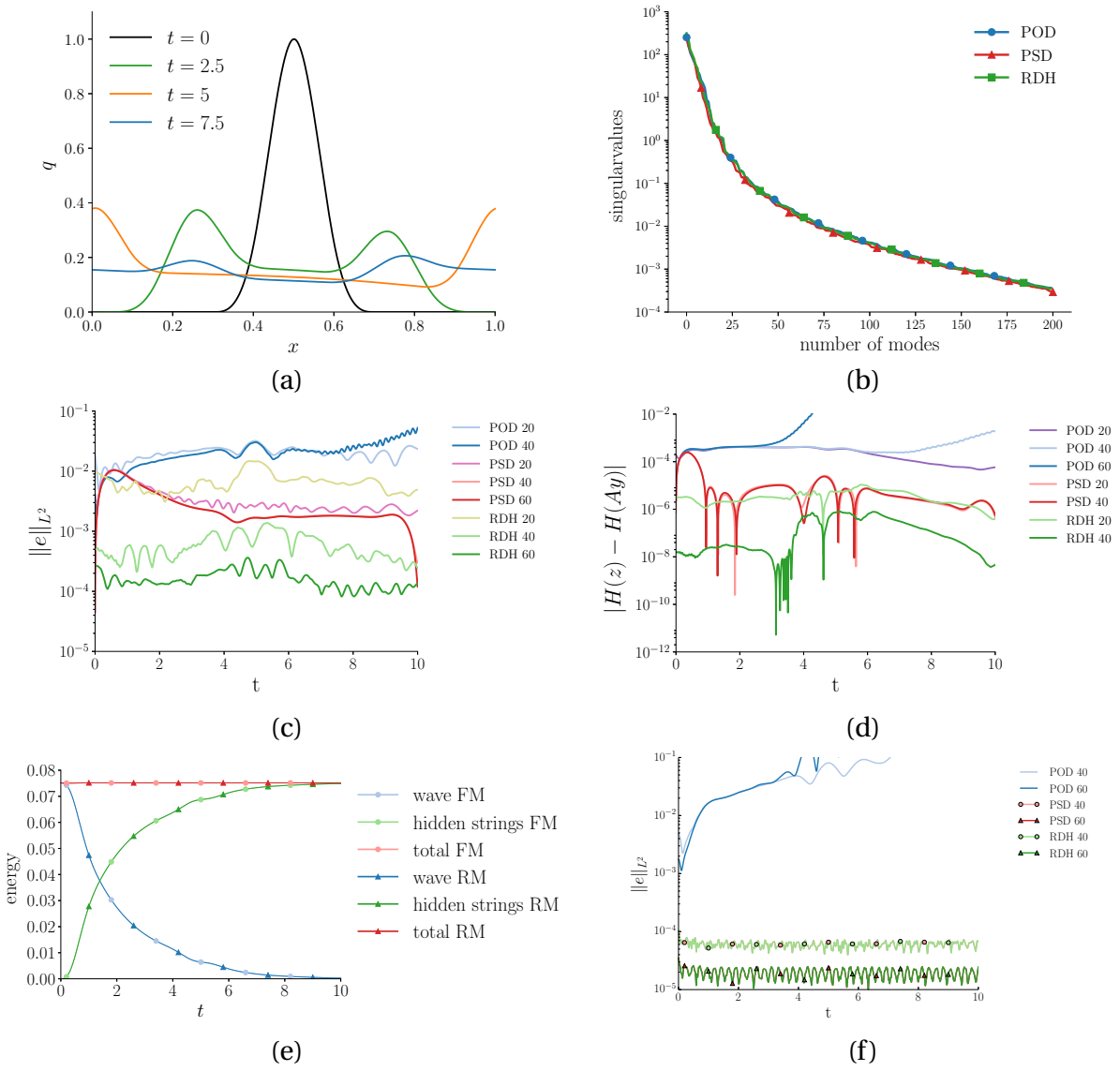


Figure 6.1 – (a) The solution to the original dissipative wave equation (6.25), (b) The decay of the singular values for the POD, the PSD, and the RDH methods, (c) The L^2-error for the different methods, (d) Evolution of error in the Hamiltonian for different methods, (e) Energy preservation of the Hamiltonian extension for the original and the reduced system. "FM" and "RM" refer to the full model and the reduced model, respectively. (f) The L^2-error between the solution to the reduced system and the full system in a near-zero dissipation regime.

In Figure 6.1.(e) we show the transfer of the energy from the TDD system to the hidden strings, for the full system and the RDH reduced system. We notice that the RDH method preserves the total energy of the extended Hamiltonian system. Furthermore, the transfer of energy to the hidden strings in the full model is correctly translated in the reduced system.

The second numerical experiment is the dissipative wave equation (6.25) in a near-zero dissipation regime. The numerical setting is taken to be identical to the previous numerical experiment, but with the difference that $r_i = 10^{-5}$, for $i = 1, \dots, N_{\Delta x}$. Figure 6.1.(f) shows the L^2-error between the solution to the reduced system and the full system, for the POD, the PSD, and the RDH methods. We notice that the POD does not yield a stable reduced system as the symplectic structure is lost via model reduction. Furthermore we notice that error for the PSD and the RDH coincide as the two methods become identical as $\|\chi\|_\infty \to 0$. Note that in this case the basis for the RDH method and PSD method are both generated using the cotangent lift method in order to show the convergence.

6.3.2 The sine-Gordon equation

Consider the one-dimensional dissipative nonlinear wave equation

$$\begin{cases} q_t(t,x) = p(t,x), \\ p_t(t,x) = q_{xx}(t,x) - \sin(q) - r(x)p(t,x), \\ q(0,x) = q_0(x), \quad q(t,0) = a, \quad q(t,L) = b, \\ p(0,x) = p_0(x). \end{cases} \tag{6.31}$$

defined on a domain of length L, which is known as the sine-Gordon equation. In the absence of dissipation, $r(x) = 0$, the *kink* solution to (6.31) is given as

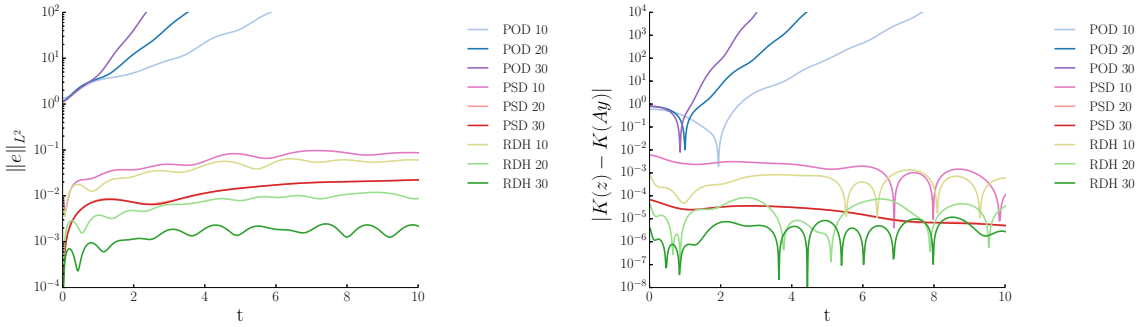
$$q(t,x) = 4 \arctan\left(\exp\left(\frac{(x - x_0 - vt)}{\sqrt{1-v^2}}\right)\right), \tag{6.32}$$

where $|v| < 1$ is the wave speed. In the presence of dissipation, where $r(x) \geq 0$, the traveling wave de-accelerates and stops. The TDD formulation for (6.31) takes the form

$$\dot{z} = \mathbb{J}_{2n} K^T f(t) + \mathbb{J}_{2n} g(z) - \mathbb{J}_{2n} z_{\text{bd}}, \quad f(t,z) + R \int_0^t f(s,z)\, ds = Kz. \tag{6.33}$$

where z, K, R and f are defined similar to (6.28), with $r_\Delta = rI_n$, and

$$g(z) = (\sin(q_1), \dots, \sin(q_{N_{\Delta x}}), 0, \dots, 0)^T, \tag{6.34}$$



(a) (b)

Figure 6.2 – (a) The L^2-error for the different methods, (a) Evolution of error in the kinetic energy for different methods.

and z_{bd} is the term corresponding to the Dirichlet boundary condition. Note that the extended Hamiltonian H_{ex} takes the form

$$H_{\text{ex}}(z,\phi,\theta) = \frac{1}{2}\left(\|Kz - \phi(t,0)\|_2^2 + \|G(z)\|_2^2 + \|\theta(t)\|_{\mathcal{H}^{2n}}^2 + \|\partial_x\phi(t)\|_{\mathcal{H}^{2n}}^2\right), \quad (6.35)$$

where $G(z)$ is a potential for $g(z)$ given as $G(q_i) = 1 - \sin(q_i)$, for $i = 1,\ldots,N_{\Delta_x}$. System parameters are summarized below

Domain length	$L = 50$
No. grid points	$N = 500$
Space discretization size	$\Delta x = L/N$
Time discretization size	$\Delta t = 0.02$
Wave speed	$v = 0.5$
Boundary conditions	$a = 0, b = 1$
Dissipation coefficient	$r = 0.1$

The RDH reduced system is constructed following Algorithm 6.1. To reduce the complexity of the nonlinear term, we used the symplectic DEIM discussed in Section 4.5. The performance of the method is then compared to the SVD based method introduced in Section 4.2.1 (here referred to as the "PSD") and the POD. To reduce the complexity of the nonlinear terms we compare the symplectic DEIM with the method proposed in [81] and the classical DEIM in Section 3.5

Figure 6.2.(a) shows the L^2-error between the full system and the RDH, the PSD, and the POD methods. Although the Hamiltonian system of the sine-Gordon equation is nonlinear, the errors for the different methods show a similar behavior as those in Section (6.3.1). We observe that the POD does not yield a stable reduced system

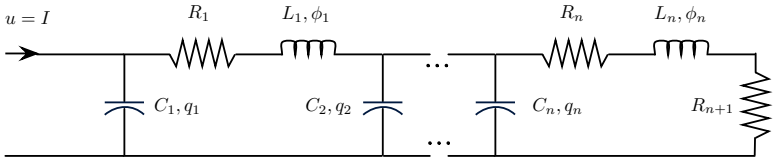


Figure 6.3 – n-dimensional ladder network

while the symplectic methods provide a high accuracy. Furthermore, we notice that enriching the PSD basis does not significantly enhance the accuracy of the method.

The evolution of error in the kinetic energy $K(p) = \|p\|_2^2/2$ is illustrated in Figure 6.2.(b). We see that the POD does not conserve the evolution of the kinetic energy. The RDH method conserves the kinetic energy of the system with a higher accuracy than the PSD method. Furthermore, the accuracy of the RDH method is better scaled under enrichment of the reduced basis, compared to the PSD method.

It is observed in Figure 1 that the symplectic treatment of the nonlinear terms is essential in correct model reduction of Hamiltonian systems. In addition, the SDEIM can be combined with the RDH method to construct a reduced Hamiltonian system that can be integrated using a symplectic integrator. Thus, the combination preserves the system energy and the symplectic symmetry of Hamiltonian systems.

6.3.3 Port-Hamiltonian Systems

Port-Hamiltonian systems are popular in network modeling and electrical engineering. In the framework of port-Hamiltonian modelling, energy conservation and Hamiltonian structure is the fundamental principle of the dynamics of the system. Ports in the system network then allow the exchange of energy with the environment in the form of sources, capacitors, and dissipations [106]. Port-Hamiltonian systems can be viewed as a forced and dissipative Hamiltonian system.

Consider the n-dimensional linear ladder network in Figure 6.3. Here C_i, L_i and R_i, $i = 1, \ldots, n$, are the capacitance, inductance and resistance of the corresponding capacitors, inductors, and resistors, respectively, and R_{n+1} is the load capacitor. The port-Hamiltonian model of the linear ladder network takes the form

$$\dot{x} = (J_{2n} - R)Q^T Q x + u. \tag{6.36}$$

Here $x = (c_1, \phi_1, \ldots, c_n, \phi_n)^T$ where c_i and ϕ_i, for $i = 1, \ldots, n$, are the charge and the flux of C_i and L_i respectively. Q and R are given as

$$Q = \text{diag}(C_1^{-1}, L_1^{-1}, \ldots, C_n^{-n}, L_n^{-n}), \quad R = \text{diag}(0, R_1, \ldots, 0, R_n + R_{n+1}), \tag{6.37}$$

$u = (1, 0, \ldots, 0)^T$ is a constant input current and J_{2n} is a skew-symmetric $2n \times 2n$ matrix with -1 and 1 on the superdiagonal and subdiagonal, respectively.

The energy associated with a port-Hamiltonian system of the form (6.36) at time t, is given as $H(x(t)) = \frac{1}{2}x^T Q^T Q x$. Since J_{2n} is skew symmetric we have that $\frac{d}{dt}H(x) = u^T Q^T Q x - x^T Q^T Q R Q^T Q x \leq u^T Q^T Q x$ which is referred to as the *passivity* of the system (6.36) [105, 110].

Since J_{2n} is full rank, one can always find a coordinate transformation $x = T\tilde{x}$ such that $T^{-1}J_{2n}T^{-T} = \mathbb{J}_{2n}$. The dissipative Hamiltonian formulation of (6.36) takes the form

$$\dot{\tilde{x}} = \mathbb{J}_{2n}\tilde{Q}^T\tilde{Q}\tilde{x} - \tilde{R}x + \tilde{u}, \tag{6.38}$$

where $\tilde{Q} = QT$, $\tilde{R} = T^{-1}RT^{-T}Q^TQ$ and $\tilde{u} = T^{-1}u$. Note that in this case, $\tilde{R}$ is symmetric and semi-positive definite since T is orthogonal and R is diagonal. The TDD formulation of (6.38) takes the form

$$\dot{\tilde{x}} = \mathbb{J}_{2n}\tilde{Q}^T f(t, \tilde{x}) + \tilde{u}, \quad f(t, \tilde{x}) + \tilde{R}\int_0^t f(t, \tilde{x}) = \tilde{Q}\tilde{x}. \tag{6.39}$$

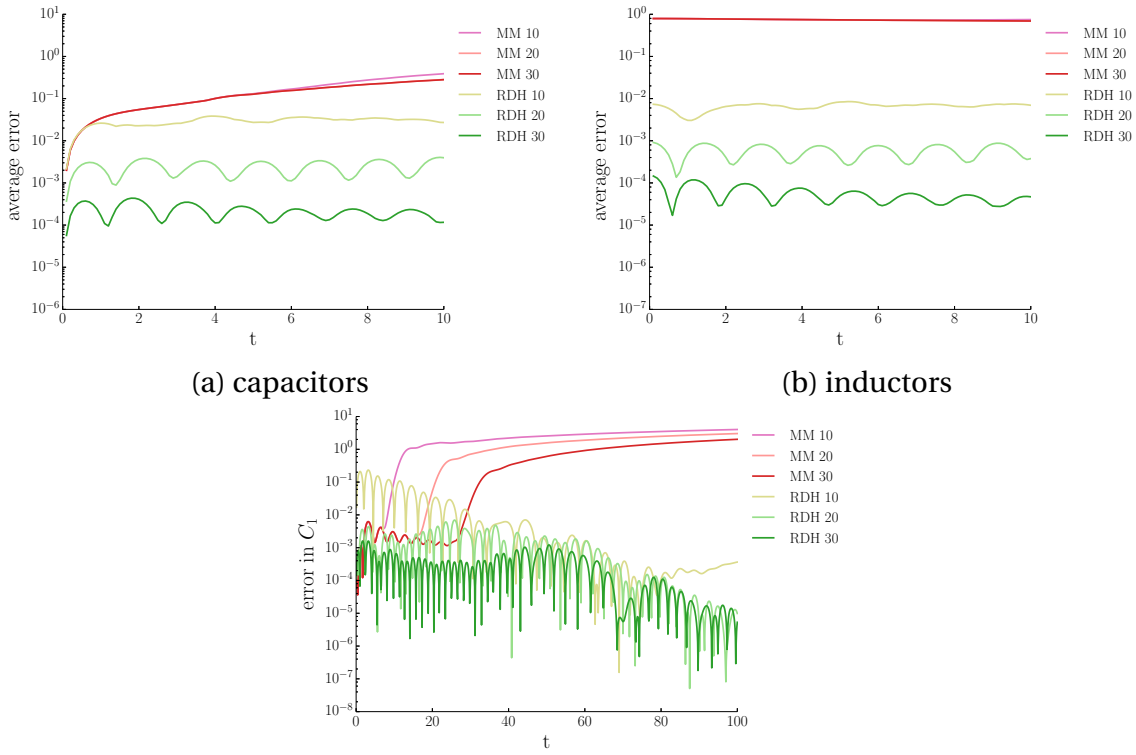
The extended Hamiltonian formulation Equations (6.8a) to (6.8b) with a quadratic Hamiltonian H_{ex} can be carried out following Section 6.1. We note that due to the input $\tilde{u}$, the Hamiltonian H_{ex} is time dependent. In fact $\frac{d}{dt}H_{\text{ex}} = \tilde{u}^T\tilde{Q}^T\tilde{Q}\tilde{x}$. If we define $\overset{\circ}{H}_{\text{ex}} : \mathbb{R}^{2n} \times \mathcal{H}^{2n} \times \mathbb{R}^2 \to \mathbb{R}$ as

$$\overset{\circ}{H}_{\text{ex}}(\tilde{x}, \phi, \theta, t, e) = H_{\text{ex}}(\tilde{x}, \phi, \theta, t) + e, \quad \dot{e} = -\partial_t H_{\text{ex}}, \tag{6.40}$$

it is easily checked that $\frac{d}{dt}\overset{\circ}{H}_{\text{ex}} = 0$ [51]. However for the time integration of the Hamiltonian system related to $\overset{\circ}{H}_{\text{ex}}$ we can apply a symplectic integrator directly on (6.40), since the evolution of $\tilde{x}$, ϕ and θ does not explicitly depend on e. Thus, the passivity of (6.36) will be preserved through a symplectic time integration of (6.39).

Using an ortho-symplectic reduced basis A, the reduced dissipative Hamiltonian method can be applied to (6.39) to construct a reduced system of the form (6.20a) to (6.20b) together with the extended Hamiltonian $\tilde{H}_{\text{ex}}$. We note that the method is also passivity preserving since $\frac{d}{dt}\tilde{H}_{\text{ex}} = (A^+\tilde{u})^T A^T\tilde{Q}^T\tilde{Q}Ay$. Furthermore, the dissipative Hamiltonian structure of the reduced system indicates that the reduced system also carries a port-Hamiltonian structure.

We consider a 100-dimensional ($n = 50$) port-Hamiltonian system for the ladder network discussed above. We take $C_i = 1$, $L_i = 1$, $R_i = 0.2$ for $i = 1, \ldots, 50$, and $R_{51} = 0.4$. We construct the RDH reduced system following Algorithm 6.1.



(a) capacitors

(b) inductors

(c) charge in C_1

Figure 6.4 – Error between the full model and the reduced model obtained by the reduced dissipative Hamiltonian method "RDH" and the moment matching method "MM". (a) The average temporal error of charge in capacitors. (b) the average temporal error of the flux in inductors. (c) The error in C_1.

The solution of the RDH method is compared to the main results of [84], where a passivity-preserving model reduction is developed using a moment matching method at infinity. The charge in C_1 is chosen to be the single out put for the moment matching method.

Reduced bases of size $2k = 10$, $2k = 20$ and $2k = 30$ are constructed with the RDH and the moment matching method. Figure 6.4.(c) shows the error in the charge in C_1 for the two methods. We observe that although the moment matching method is bounded over long-time integration, the RDH method provides a significantly more accurate solution. In the moment matching method, the passivity of the reduced system implies that the energy of the system will be bounded by the input energy. However, there is no guarantee that the dissipation of energy in the reduced system mimics the one of the original system. On the other hand, the RDH method allows a correct dissipation of energy through the hidden strings and the symplectic time integration in the RDH method guarantees that the total energy is preserved.

Over short-time integration, we notice that the moment matching method with 10

modes provides a more accurate solution than the RDH with 10 modes. Furthermore, the moment matching method with 20 and 30 modes provide a comparable accuracy to the RDH method with 20 and 30 modes. However, the RDH method maintains the high accuracy during long-time integration, while the moment matching method loses up to 3 orders of magnitude in the accuracy, independent of the number of modes.

Figure 6.4.(a) and Figure 6.4.(b) show the average temporal error in the charge and flux of the capacitors and inductors, respectively. The RDH method provides a significantly better accuracy compared to the moment matching method. This is because the charge of C_1 is specified as the output of interest in the moment matching method and so it is expected that that method provides low accuracy for computing other outputs. On the other hand, the RDH method not only provides high accuracy in computing the charge for C_1 but also high accuracy for all components of the system.

6.4 Conclusion

In this chapter we presented the reduced dissipative Hamiltonian method. The method preserves the symplectic structure of dissipative Hamiltonian systems and guarantees the correct dissipation of energy through time integration. The RDH method couples the reduced system with a canonical heat bath such that the reduced system forms a closed system.

The main advantage of the RDH method compared to the existing methods is that it enables the reduced system to be integrated using a symplectic integrator which naturally preserves the Hamiltonian structure and the symplectic symmetry of the Hamiltonian systems. Applying a symplectic integrator to a non-conservative system or using a non-symplectic integrator for the reduced system can cause accumulation of local errors or wrong qualitative solution over long-time integration, respectively.

The numerical simulations illustrate that the RDH method preserves the system energy with significantly higher accuracy than other methods. Furthermore, it is shown that the hidden strings assure that the dissipation of energy in the reduce system mimics the dissipation of energy in the full system. This ensures that the local error do not accumulate over long-time integration.

7 Conservative Model Order Reduction of Fluid Flow

In Chapters 4 to 6 it is discussed how MOR can be modified to ensure conservation of symmetries and invariants in the context of Hamiltonian system. A reduced Hamiltonian, as an approximation of the Hamiltonian for the original system, is introduced for the reduced system. Conservation of a skew-symmetric form can then ensures that correct evolution of the reduced system and conservation of the reduced Hamiltonian. In this chapter, we apply the same principle to construct a MOR technique for the fluid flow that conserves energy of the flow.

Large scale simulations of fluid flows arise in a wide range of disciplines and industries. Therefore, MOR of fluid flows, specially when advective terms are dominant, is important. It is well known that conservation of the energy, specially kinetic energy, is essential for a qualitatively correct numerical integration of fluid flows. Conventional model reduction techniques often violate conservation of mass, momentum [22], or energy in fluid flows which result in an unstable reduced system, in particular for long time-integration. In [7] an entropy stable model reduction method for linear compressible flows is presented by considering an entropy-stable formulation of linearized compressible flows. Furthermore, a conservative model reduction for finite-volume models is presented in [22] that that conserves any quantity conserved by the finite-volume scheme. This method finds a reduced linear subspace that ensures conservation of quantities by solving an optimization problem with, generally nonlinear, equality constraints that has to be solved in each time instance.

The method discussed in [22] conserves the mass and momentum for a finite volume discretization scheme. However, a method that also conserves the energy of the fluid flow is not known to the authors.

Skew-symmetric formulation of fluid flows constructs a skew-symmetric differential operator, acting on the momentum vector field, that ensures conservation of quadratic invariants, such as energy. Combined wirh centered time and space discretization schemes, typically a finite differences discretization method, they recover

time-symmetries of a fluid at the discrete level. Such discretization schemes are studies comprehensively over the past few decades and can be found in the works of [75, 76, 31, 101, 88] and the references therein.

This chapter discusses how to preserve skew-symmetry of the differential operators at the level of the reduced system. This results in conservation of quadratic invariants. As the reduced system is constructed using a proper orthogonal decomposition (POD) basis, the proposed method avoids solving nonlinear optimization problems at each time instance. This saves significant costs at the online stage, compared to other conservative model reduction methods for fluid flow. Furthermore, we show that the reduced system, as a system of coupled differential equations, contains quadratic invariants and an associated energy which approximates the energy of the high-fidelity system. Therefore, a proper time stepping scheme preserves the reduced representation of the energy, and therefore, the loss in energy due to model reduction remains constant in time.Furthermore, we demonstrate, through numerical experiments, that a quasi-skew-symmetric form of fluid flow, i.e. a formulation where only spacial differential operators are in a skew-symmetric form, offer remarkable stability properties in terms of MOR. This allows an explicit time-integration to be utilized while recovering robustness of skew-symmetric forms at the reduced level.

The rest of this chapter is organized as follows. We discuss skew-symmetric and conservatives methods for compressible and incompressible fluid flows in Section 7.1. Conservative and energy-preserving model reduction of fluid flows is discussed in Section 7.2. We evaluate the performance of the method through numerical simulations of incompressible and compressible fluid flow in Section 7.3. We also apply the method to construct a reduced system for the continuous variable resonance combustor, a one dimensional reaction-diffusion model for a rocket engine. Finally, we present conclusive remarks in Section 7.4.

7.1 Skew Symmetric and Centered Schemes for Fluid Flows

In this section we summarize the conservation properties of skew-symmetric forms and discretization schemes, following, closely, the works of [75, 76, 101, 88].

7.1.1 Conservation Laws

In the context of fluid flows, transport of conserved quantities, can be expressed as

$$\frac{\partial}{\partial t}\rho\varphi + \nabla\cdot(\rho u\varphi) = \nabla\cdot F_\varphi \quad \text{defined in} \quad \Omega\subset\mathbb{R}^d. \tag{7.1}$$

Here, $d = 1, 2$ or 3, $\rho : \Omega \to \mathbb{R}$ is the density, $u : \Omega \to \mathbb{R}^d$ is the velocity vector field, φ is a measured scalar quantity of the flow, and F_φ is the flux function associated to φ. Integration of (7.1) over Ω yields

$$\frac{d}{dt}\int_\Omega \rho\varphi \, dx = \int_{\partial\Omega} (F_\varphi - \rho u\varphi) \cdot \hat{n} \, ds, \tag{7.2}$$

where $\partial\Omega$ is the boundary of Ω, and $\hat{n}$ is the unit outward normal vector to $\partial\Omega$. This means that the quantity $(\rho\varphi)$ is explicitly conserved over control volumes. Therefore, (7.2) is referred to as the *conservative form* and the convective term in (7.1) is referred to as the *divergence form*. However, using the *continuity equation*

$$\frac{\partial}{\partial t}\rho + \nabla \cdot (\rho u) = 0, \tag{7.3}$$

we can rewrite (7.1) as

$$\rho\frac{\partial}{\partial t}\varphi + (\rho u) \cdot \nabla\varphi = \nabla \cdot F_\varphi. \tag{7.4}$$

The convective term in this formulation is referred to as the *advective form*. The *skew-symmetric* form of the convective term is obtained by the arithmetic average of the divergent and the advective form:

$$\frac{1}{2}\left(\rho\frac{\partial}{\partial t}\varphi + \frac{\partial}{\partial t}(\rho\varphi)\right) + \frac{1}{2}\left((\rho u) \cdot \nabla\varphi + \nabla \cdot (\rho u\varphi)\right) = \nabla \cdot F_\varphi. \tag{7.5}$$

Multiplying (7.5) with φ yields

$$\frac{1}{2}\left(\rho\varphi\frac{\partial}{\partial t}\varphi + \varphi\frac{\partial}{\partial t}(\rho\varphi)\right) + \frac{1}{2}\left((\rho u\varphi) \cdot \nabla\varphi + \varphi\nabla \cdot (\rho u\varphi)\right) = \varphi\nabla \cdot F_\varphi. \tag{7.6}$$

Using the product rule, we recover

$$\frac{\partial}{\partial t}\rho\varphi^2 + \nabla \cdot (\rho u\varphi^2) = \varphi\nabla \cdot F_\varphi. \tag{7.7}$$

Therefore, φ^2 is a conserved quantity for a flux-free φ. Since the divergence, the advective and the skew-symmetric forms are identical at the continuous level, φ^2 is a conserved quantity for all forms. However, the equivalence of these forms is not preserved through a general discretization scheme and we can not expect φ^2 to be a conserved quantity at the discrete level. To motivate numerical advantages of the skew-symmetric form consider the operator

$$S_{\rho u}(\cdot) = \frac{1}{2}([\nabla \cdot \rho u](\cdot) + (\rho u) \cdot \nabla(\cdot)), \tag{7.8}$$

where $[\nabla \cdot \rho u](x) = \nabla \cdot (\rho u x)$. With a proper set of boundary condition, this operator is a skew-adjoint operator on L^2. Here, $[\cdot]$ indicates that the inside of the brackets act as a differential operator. This skew-adjoint property is used later to show the conservation of some quadratic quantities in (3.1). Similarly, we can define a skew-adjoint operator with respect to the time variable as

$$S_{\rho,\partial_t} = \frac{1}{2}\left(\rho\frac{\partial}{\partial t} + [\frac{\partial}{\partial t}\rho]\right). \tag{7.9}$$

Here, the subscript ∂t is to emphasize that S_{ρ,∂_t} is a differential operator with respect to t. A proper time and space discretization of $S_{\rho u}$ and S_{ρ,∂_t} can preserve the skewness property.

Numerical time integration of (7.5) can be challenging since the time differentiation of different variables is present. Following [75], we have

$$\begin{aligned}\frac{1}{2}\left(\frac{\partial}{\partial t}(\rho\varphi) + \rho\frac{\partial}{\partial t}\varphi\right) = \left(\frac{\partial}{\partial t}(\rho\varphi) - \frac{\varphi}{2}\frac{\partial}{\partial t}\rho\right) = \left(\rho\frac{\partial}{\partial t}\varphi + \frac{\varphi}{2}\frac{\partial}{\partial t}\rho\right)\\ = \sqrt{\rho}\frac{\partial}{\partial t}(\sqrt{\rho}\varphi),\end{aligned} \tag{7.10}$$

where the product rule is used in the last step. Substituting this into (7.5) yields

$$\sqrt{\rho}\frac{\partial}{\partial t}(\sqrt{\rho}\varphi) + S_{\rho u}(\varphi) = \nabla \cdot F_\varphi. \tag{7.11}$$

Time integration of this form is presented in [75, 88]. Note that one can also generate a quasi-skew-symmetric form [17, 77] of (7.1) as

$$\frac{\partial}{\partial t}(\rho\varphi) + \frac{1}{2}\left(\nabla \cdot (\rho u\varphi) + \rho u \cdot \nabla\varphi + \varphi\nabla \cdot (\rho u)\right) = \nabla \cdot F_\varphi. \tag{7.12}$$

Here, we substitute the convective term in (7.1) with

$$\nabla \cdot (\rho u\varphi) = \frac{1}{2}(2\nabla \cdot (\rho u\varphi)) = \frac{1}{2}\left(\nabla \cdot (\rho u\varphi) + \rho u \cdot \nabla\varphi + \varphi\nabla \cdot (\rho u)\right). \tag{7.13}$$

Even though this is not a fully skew-symmetric form (skew-symmetric only in space), the numerical stability of this form is significantly better than the divergence and advective form [75, 17, 77]. Note that this quasi-skew-symmetric form is identical to the skew-symmetric form in the incompressible limit.

7.1.2 Incompressible Fluid

Consider the governing equations of an incompressible fluid with skew-symmetric convective term:

$$\begin{cases} \nabla \cdot u = 0, \\ \dfrac{\partial}{\partial t} u + S_u(u) + \nabla p = \nabla \cdot \tau, \end{cases} \tag{7.14}$$

defined on Ω. Here, $p : \Omega \to \mathbb{R}^+$ is the pressure, $\tau : \Omega \to \mathbb{R}^{d\times d}$ is the viscous stress tensor, and $S_u = \frac{1}{2}([\nabla \cdot u] + u \cdot \nabla)$. It is straight forward to check

$$\frac{d}{dt}K + \nabla \cdot (Ku) + \nabla \cdot (pu) = \nabla \cdot (\tau u) - (\tau \nabla) \cdot u, \tag{7.15}$$

where $K = \frac{1}{2}\sum_{i=1}^{d} u_i^2$ is the kinetic energy and we used

$$u \cdot S_u(u) = \nabla \cdot (Ku). \tag{7.16}$$

The only non-conservative term in (7.15) is $-(\tau\nabla)\cdot u$, which corresponds to dissipation of kinetic energy. Therefore, in the absence of the viscous terms, K is a conserved quantity of the system, and $\frac{d}{dt}\int_\Omega K\, dx < 0$ when $\tau \neq 0$. Note that as long as $\nabla \cdot u = 0$, as discussed in Section 7.1.1, the divergence, the convective, and the skew-symmetric forms are identical for the incompressible fluid equation. Thus, kinetic energy is conserved for all forms. However, for a general discretization scheme, these forms are not identical and often conservation of kinetic energy (in the discrete sense) may be violated.

A skew-symmetric discretization of (7.14) is a centered scheme that exploits the skew-adjoint property of S_u, and ensures conservation of kinetic energy at the discrete level. We uniformly discretize Ω into N points and denote by $\mathbf{u} \in \mathbb{R}^{N\times d}$, $\mathbf{p} \in \mathbb{R}^N$, and $T \in R^{N\times d\times d}$ the discrete representation of u, p, and τ, respectively. Let D_j be the centered finite difference scheme for $\partial/\partial x_j$, and for $j = 1, \ldots, d$. The momentum equation in (7.14) is discretized as

$$\frac{d}{dt}\mathbf{u}_i + S_{\mathbf{u}}\mathbf{u}_i + D_i\mathbf{p} = \sum_{j=1}^{d} D_j T_{ij}, \quad i = 1, \ldots, d, \tag{7.17}$$

where $S_{\mathbf{u}}$ is the discretization of S_u given by

$$S_{\mathbf{u}} = \sum_{j=1}^{d} D_j U_j + U_j D_j, \tag{7.18}$$

and U_i contains components of $\mathbf{u}_i$ on its diagonal. We require D_i to satisfy

1. $D_i = -D_i^T$

2. $D_i\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ and $\mathbf{0}$ are vectors of ones and zeros, respectively.

Conditions 1 and 2 yield

$$S_{\mathbf{u}} = -S_{\mathbf{u}}^T, \quad \mathbf{1}^T S_{\mathbf{u}} \mathbf{u}_i = 0, \quad i = 1, \ldots, d. \tag{7.19}$$

Conservation of momentum in the discrete sense is expressed as

$$\frac{d}{dt} \sum_{i=1}^{d} \mathbf{1}^T \mathbf{u}_i = \sum_{i=1}^{d} \left(-\mathbf{1}^T S_{\mathbf{u}} \mathbf{u}_i - \mathbf{1}^T D_i \mathbf{p} + \sum_{j=1}^{d} \mathbf{1}^T D_j T_{ij} \right) = 0. \tag{7.20}$$

Similarly, it is verified that

$$\frac{d}{dt} \sum_{i=1}^{d} \left(\frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i \right) = - \sum_{i,j=1}^{d} T_{ij} D_j \mathbf{u}_i \leq 0. \tag{7.21}$$

Conditions 1 and 2 for D_i are easily checked for a centered finite differences scheme on a periodic domain. For other types of boundaries, e.g., wall boundary and inflow/outflow, we refer the reader to [76, 31] for the construction of the proper discrete centered differentiation operator. We note that the finite differences schemes are chosen here for illustration purposes. It is easily checked that any discrete differentiation operator that satisfies discrete integration by parts, e.g. summation by part (SBP) methods and discontinuous Galerkin (DG) methods, also satisfies conditions 1 and 2 and can be used to construct a skew-symmetric discretization.

7.1.3 Compressible Fluid

Consider the equations governing the evolution of a compressible fluid in a skew-symmetric form in one spacial dimension

$$\begin{cases} \dfrac{\partial}{\partial t}\rho + \dfrac{\partial}{\partial x}(\rho u) = 0, \\ S_{\rho,\partial_t}(u) + S_{\rho u}(u) + \dfrac{\partial}{\partial x} p = \dfrac{\partial}{\partial x}\tau, \\ \dfrac{\partial}{\partial t}\rho E + \dfrac{\partial}{\partial x}(uE + up) = \dfrac{\partial}{\partial x}(u\tau - \phi). \end{cases} \tag{7.22}$$

Here $E = e + u^2/2$ is the total energy per unit mass, with $e = p/\rho(\gamma - 1)$ being the internal energy, γ the adiabatic gas index, and $\phi = -\lambda \frac{\partial T}{\partial x}$ is the heat flux, with λ as the heat conductivity. The remaining variables are the same as those discussed in

Section 7.1.2. Following [88], the evolution of the momentum equation is

$$\begin{aligned}\frac{\partial}{\partial t}(\frac{\rho u^2}{2})+\frac{\partial}{\partial x}(\rho u\frac{u^2}{2}) &= \frac{1}{2}u(\frac{d}{dt}\rho u+\rho\frac{d}{dt}u)+\frac{1}{2}u([\frac{\partial}{\partial x}\rho u]u+\rho u\frac{\partial}{\partial x}u)\\ &= -u\frac{\partial}{\partial x}p+u\frac{\partial}{\partial x}\tau.\end{aligned} \tag{7.23}$$

Substituting this into the energy equation in (7.22), while assuming a constant adiabatic index, yields

$$\frac{1}{\gamma-1}\frac{d}{dt}p+\frac{\gamma}{\gamma-1}\frac{\partial}{\partial x}up-u\frac{\partial}{\partial x}(p)=-u\frac{\partial}{\partial x}\tau+\frac{\partial}{\partial x}(u\tau-\phi). \tag{7.24}$$

We discretize the real line, uniformly, into N grid points and denote by $\mathbf{r},\mathbf{u},\mathbf{p}\in\mathbb{R}^N$, the discrete representations of ρ, u, and p, respectively. Using the matrix differentiation operator $D\in\mathbb{R}^{N\times N}$ (we omit the subscript "i" for the one dimensional case), introduced in Section 7.1.2, we define the skew-symmetric matrix operator $S_{\mathbf{ru}}=\frac{1}{2}(DUR+RUD)$, where R is the matrix that contains r in its diagonal. Semi-discrete expression of (7.22) and (7.24) takes the form

$$\begin{cases}\dfrac{d}{dt}\mathbf{r}+DU\mathbf{r}=0,\\ S_{\mathbf{r},\partial_t}(\mathbf{u})+S_{\mathbf{ru}}\mathbf{u}+D\mathbf{p}=DT,\\ \dfrac{1}{\gamma-1}\dfrac{d}{dt}\mathbf{p}+\dfrac{\gamma}{\gamma-1}DU\mathbf{p}-UD\mathbf{p}=-UDT+D(UT-\phi).\end{cases} \tag{7.25}$$

Recalling conditions 1 and 2 for D, discussed in Section 7.1.2, it is easily verified that

$$S_{\mathbf{ru}}^T=-S_{\mathbf{ru}},\quad \mathbf{1}^TS_{\mathbf{ru}}\mathbf{u}=-\mathbf{u}^TDU\mathbf{r}. \tag{7.26}$$

Conservation of mass is expressed as

$$\frac{d}{dt}(\mathbf{1}^T\mathbf{r})=-\mathbf{1}^TDR\mathbf{u}=0. \tag{7.27}$$

Furthermore, we recover conservation of momentum in the discrete sense as

$$\begin{aligned}\frac{d}{dt}(\mathbf{r}^T\mathbf{u}) &= \frac{1}{2}\frac{d}{dt}(\mathbf{r}^T\mathbf{u})+\frac{1}{2}\left(\mathbf{r}^T\frac{d}{dt}\mathbf{u}+\mathbf{u}^T\frac{d}{dt}\mathbf{r}\right)\\ &= \frac{1}{2}u^T\frac{d}{dt}\mathbf{r}+\mathbf{1}^TS_{\mathbf{r},\partial_t}(\mathbf{u})\\ &= -\frac{1}{2}\mathbf{u}^TDU\mathbf{r}-\frac{1}{2}\mathbf{1}^TS_{\mathbf{ru}}\mathbf{u}-\mathbf{1}^TD\mathbf{p}+\mathbf{1}^TDT=0.\end{aligned} \tag{7.28}$$

Here we used (7.26) and the mass and the momentum equation in (3.22). Similarly, for conservation of the total energy, we have

$$\frac{d}{dt}\left(\frac{1}{\gamma-1}\mathbf{1}^T\mathbf{p}+\frac{1}{2}(R\mathbf{u})^T\mathbf{u}\right)=\frac{d}{dt}\left(\frac{1}{\gamma-1}\mathbf{1}^T\mathbf{p}\right)+\frac{1}{2}\mathbf{u}^T S_{\mathbf{r},\partial_t}(\mathbf{u})=0. \tag{7.29}$$

In addition to the conservation of the total energy, the skew-symmetric form of (7.25) also conserves the evolutions of the kinetic energy:

$$\begin{aligned}\frac{d}{dt}(\frac{1}{2}\mathbf{u}^T R\mathbf{u})=\frac{1}{2}\mathbf{u}^T S_{\mathbf{r},\partial_t}(\mathbf{u})&=-\mathbf{u}^T S_{\mathbf{ru}}\mathbf{u}+\mathbf{u}^T Dp+\mathbf{u}^T DT\\&=\mathbf{u}^T Dp+\mathbf{u}^T DT.\end{aligned} \tag{7.30}$$

Here, we used the skew-symmetry of $S_{\mathbf{ru}}$. Therefore, only the pressure and the viscous terms contribute to a change in the kinetic energy.

We point out that there are other methods to obtain a skew-symmetric form for (7.22), that result in the conservation of other quantities. An entropy preserving skew-symmetric form can be found in [98]. Furthermore, a fully quasi-skew-symmetric form for (7.22), where all quadratic fluxes are in a skew-symmetric form, is shown to minimize aliasing errors [54, 53]

7.1.4 Time integration

Following [88, 75] we can construct a fully discrete second order accurate scheme for (7.1.3) as

$$\begin{cases}\frac{1}{2}\sqrt{\mathbf{r}}^{n+1/2}\frac{\sqrt{\mathbf{r}}^{n+1}-\sqrt{\mathbf{r}}^{n}}{\Delta t}+DU^{n+1/2}\mathbf{r}^n=0,\\ \sqrt{\mathbf{r}}^{n+1/2}\frac{\sqrt{\mathbf{R}}^{n+1}u^{n+1}-\sqrt{\mathbf{R}}^{n}u^{n}}{\Delta t}+S_{\mathbf{r}^n\mathbf{u}^n}\mathbf{u}_\alpha^{n+1/2}+D\mathbf{p}^n=DT^n,\\ \frac{1}{\gamma-1}\frac{\mathbf{p}^{n+1}-\mathbf{p}^n}{\Delta t}+\frac{\gamma}{\gamma-1}DU^n\mathbf{p}-U^nD\mathbf{p}=-U^nDT^n+D(U^nT^n-\phi^n).\end{cases} \tag{7.31}$$

Here, Δt is the time step, superscript n denotes evaluating at $t=n\Delta t$, superscript $n+1/2$ denotes the arithmetic average of a variable evaluated at $t=n\Delta t$ and $t=(n+1)\Delta t$, the square root sign denotes element-wise application of square root, and

$$\mathbf{u}_\alpha^{n+1/2}=\frac{\sqrt{\mathbf{R}}^{n+1}\mathbf{u}^{n+1}+\sqrt{\mathbf{R}}^{n}\mathbf{u}^{n}}{2\sqrt{\mathbf{r}}^{n+1/2}}. \tag{7.32}$$

As discussed in [88], this time discretization scheme preserves the symmetries expressed in (7.21), (7.28), (7.29), and (7.30). In the incompressible case, the method reduces to the implicit mid-point scheme [51]. For further information see [88, 75].

7.2 Model Order Reduction of Fluid Flow

A straight-forward model reduction of (7.14) and (7.22) does not, in general, preserve symmetries and conservation laws, presented in Section 7.1. In this section we discuss how to exploit the discrete skew-symmetric structure of (7.17) and (7.25) to recover conservation of mass, momentum, and energy at the level of the reduced system.

Let $V_{\mathbf{r}}$, $V_{\mathbf{ru}}$, and $V_{\mathbf{u}_i}$ be the reduced bases for the snapshots of $\mathbf{r}$, $R\mathbf{u}$, and $\mathbf{u}_i$, respectively. For the one dimensional case, the subscript "i" is omitted and for an incompressible fluid, $V_{\mathbf{r}}$ and $V_{\mathbf{ru}}$ are not computed. For the purpose of simplicity, we assume that all bases have the size k. We seek to project $S_{\mathbf{u}}$ and $S_{\mathbf{ru}}$ onto the reduced space, such that the projection preserves the skew-symmetric property. The projected operators, using a Galerkin projection, read

$$S^r_{\mathbf{u}} = V^T_{\mathbf{u}_i} S_{\mathbf{u}} V_{\mathbf{u}_i}, \quad i = 1, \dots, d, \tag{7.33}$$

and

$$S^r_{\mathbf{r},\partial_t} = V^T_{\mathbf{ru}} S_{\mathbf{r},\partial_t} V_{\mathbf{u}}, \quad S^r_{\mathbf{ru}} = V^T_{\mathbf{ru}} S_{\mathbf{ru}} V_{\mathbf{u}}. \tag{7.34}$$

Note that $S^r_{\mathbf{r},\partial_t}$ is not computed explicitly. It is clear that $S^r_{\mathbf{u}}$ is already in a skew-symmetric form. On the other hand, $S^r_{\mathbf{r},\partial_t}$ and $S^r_{\mathbf{ru}}$ are not, in general, skew-adjoint and skew-symmetric, respectively. This can be ensured, however, by requiring $V_{\mathbf{ru}} = V_{\mathbf{u}}$. We denote such a basis by $V_{\mathbf{ru},\mathbf{u}}$.Using (7.33) and (7.34), a Galerkin projection of the momentum equation in (7.17) and the governing equations for a compressible fluid in (7.25) take the form

$$\frac{d}{dt}\mathbf{u}^r{}_i + S^r_{\mathbf{u}}\mathbf{u}^r_i + V^T_{\mathbf{u}_i} D_i \mathbf{p} = \sum_{j=1}^{d} V^T_{k3,\mathbf{u}_i} D_j T_{ij} (V_{\mathbf{u}_i}\mathbf{u}^r_i), \quad i = 1, \dots, d, \tag{7.35}$$

and

$$\begin{cases} \dfrac{d}{dt}\mathbf{r}^r + \displaystyle\sum_{i=1}^{k} V^T_{\mathbf{r}} D U_i V_{\mathbf{r}} \mathbf{r}^r = 0, \\ S^r_{\mathbf{r},\partial_t} + S^r_{\mathbf{ru}}\mathbf{u}^r + V^T_{\mathbf{ru},\mathbf{u}} D V_{\mathbf{p}} \mathbf{p}^r = V^T_{\mathbf{ru},\mathbf{u}} D T, \\ \dfrac{1}{\gamma - 1}\dfrac{d}{dt}\mathbf{p}^r + \dfrac{\gamma}{\gamma - 1} V^T_{\mathbf{p}} D U V_{\mathbf{p}} \mathbf{p}^r - V^T_{\mathbf{p}} U D V_{\mathbf{p}} \mathbf{p}^r = -V^T_{\mathbf{p}} U D T + V^T_{\mathbf{p}} D (U T - \phi), \end{cases} \tag{7.36}$$

respectively. Note that in (7.36), dependency of T on $V_{\mathbf{ru},\mathbf{u}}$ is not shown for abbreviation. In (7.35) and (7.36), D_i is always multiplied from the left with a basis matrix or a diagonal matrix. Therefore, the telescoping sum, discussed in Condition 2 in Section 7.1.1, cannot be used to show conservation of mass and momentum. However, POD preserves linear properties of snapshots. To demonstrate this, let the overscript

"~" denote the representation of a reduced variable in the high-fidelity space. An approximated variable, e.g. density, can be represented as a linear combination of some snapshots as $\mathbf{r} \approx \tilde{\mathbf{r}} = \sum_{i=1}^{k} c_i \mathbf{r}_i$, for some snapshots $\mathbf{r}_i$ and some coefficients $c_i \in \mathbb{R}$, for $i = 1, \ldots, k$. Conservation of mass, evaluated by $\tilde{\mathbf{r}}$, reads

$$\frac{d}{dt}\mathbf{1}^T\tilde{\mathbf{r}} = \sum_{i=1}^{k} c_i \left(\mathbf{1}^T \frac{d}{dt}\mathbf{r}_i\right) = -\sum_{i=1}^{k} c_i \left(\mathbf{1}^T D R_i \mathbf{u}_i\right) = 0, \tag{7.37}$$

where we used the fact that $\mathbf{1}^T D = \mathbf{0}^T$. Similarly, we recover conservation of momentum

$$\begin{aligned}\frac{d}{dt}(\tilde{\mathbf{r}}^T\tilde{\mathbf{u}}) &= \frac{1}{2}\frac{d}{dt}(\tilde{\mathbf{r}}^T\tilde{\mathbf{u}}) + \frac{1}{2}\left(\tilde{\mathbf{r}}^T\frac{d}{dt}\tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T\frac{d}{dt}\tilde{\mathbf{r}}\right)\\ &= \sum_{i,j=1}^{k} d_i c_j \left(\mathbf{u}_i^T\frac{d}{dt}\mathbf{r}_j + \left(\mathbf{r}_j^T\frac{d}{dt}\mathbf{u}_i + \mathbf{u}_i^T\frac{d}{dt}\mathbf{r}_j\right)\right) = 0.\end{aligned} \tag{7.38}$$

Here, $\tilde{\mathbf{u}} = \sum_{i=1}^{k} d_i \mathbf{u}_i$, for some snapshot $\mathbf{u}_i$ and coefficients $d_i \in \mathbb{R}$. Denoting by $\{R\mathbf{u}\}^r$ the reduced representation of $R\mathbf{u}$ in basis $V_{\mathbf{ru},\mathbf{u}}$, the evolution of kinetic energy is expressed as

$$\begin{aligned}\frac{d}{dt}\left(\frac{1}{2}\tilde{\mathbf{u}}^T\tilde{R}\tilde{\mathbf{u}}\right) &= \frac{d}{dt}\left(\frac{1}{2}\mathbf{u}^{rT}V_{\mathbf{ru},\mathbf{u}}^T V_{\mathbf{ru},\mathbf{u}}\{R\mathbf{u}\}^r\right) = \frac{d}{dt}\left(\frac{1}{2}\mathbf{u}^{rT}\{R\mathbf{u}\}^r\right)\\ &= \frac{1}{2}\left(\mathbf{u}^{rT}\frac{d}{dt}\{R\mathbf{u}\}^r + \{R\mathbf{u}\}^r\frac{d}{dt}\mathbf{u}^r\right)\\ &= \frac{1}{2}\left(\mathbf{u}^{rT}V_{\mathbf{ru},\mathbf{u}}^T V_{\mathbf{ru},\mathbf{u}}\frac{d}{dt}\{R\mathbf{u}\}^r + \{R\mathbf{u}\}^r\frac{d}{dt}V_{\mathbf{ru},\mathbf{u}}^T V_{\mathbf{ru},\mathbf{u}}\mathbf{u}^r\right)\\ &= \mathbf{u}^{rT}S_{\mathbf{r},\partial_t}^r\mathbf{u}^r = \mathbf{u}^{rT}V_{\mathbf{ru},\mathbf{u}}DV_{\mathbf{p}}\mathbf{P}^r + \mathbf{u}^{rT}V_{\mathbf{ru},\mathbf{u}}^T DT.\end{aligned} \tag{7.39}$$

In the missing steps in the last line, skew-symmetry of $S_{\mathbf{ru}}^r$ is used. Note, that only the reduced pressure and the viscous term contribute to the evolution of kinetic energy. Furthermore, the quantity $\frac{1}{2}\mathbf{u}^{rT}\{R\mathbf{u}\}^r$ is the kinetic energy associated with the reduced system (7.36), approximating the kinetic energy of the high-fidelity system (7.25), and is a quadratic form with respect to the reduced variables. Conservation of kinetic energy for (7.35) follows similarly. It is straight-forward to check that

$$\frac{d}{dt}\left(\frac{1}{\gamma-1}\mathbf{1}^T\tilde{\mathbf{p}} + \frac{1}{2}\tilde{\mathbf{u}}^T\tilde{R}\tilde{\mathbf{u}}\right) = 0, \tag{7.40}$$

i.e., the total energy is conserved. We immediately recognize that $\mathbf{p}^r/(\gamma-1)$ is the internal energy of the reduced system. However, the total internal energy of (7.36) is a weighted sum, $b^T\mathbf{p}^r/(\gamma-1)$, with $b = V_{\mathbf{p}}^T\mathbf{1}$ which is an approximation of the total internal energy in (7.25). From (7.37), (7.38), (7.39), and (7.40) we conclude the following proposition.

Proposition 7.1. *The loss in the mass, momentum and energy associated with the model reduction in* (7.36) *is constant in time, and therefore, bounded.*

7.2.1 Assembling Nonlinear Terms and Time Integration

Nonlinear terms that appear in (7.35) and (7.36) are of quadratic nature. These terms can be evaluated exactly using a set of precomputed matrices as proposed in [12]. As an example, consider

$$S_{\mathbf{u}}^r = V_{\mathbf{u}}^T (DU + UD) V_{\mathbf{u}}^T. \tag{7.41}$$

We write U as a linear combination of matrices as $U = \sum_{j=1}^{k} \mathbf{u}_j^r U_j$, where $\mathbf{u}_j^r$ is the jth component of $\mathbf{u}^r$, and U_j contains the jth column of $V_{\mathbf{u}}$ on its diagonal. It follows

$$S_{\mathbf{u}}^r = \sum_{j=1}^{k} \mathbf{u}_j^r \left(V_{\mathbf{u}}^T (DU_j + U_j D) V_{\mathbf{u}}^T \right). \tag{7.42}$$

The matrices $V_{\mathbf{u}}^T (DU_j + U_j D) V_{\mathbf{u}}^T$ can be computed prior to the time integration of the reduced system. However, the form of the fully discrete system in (7.31) introduces cubic and even quartic terms. In principle, the same method can be applied to assemble the nonlinear terms. However, the number of precomputed matrices grows proportional to the order of the nonlinear term.

To accelerate assembly of the nonlinear terms we may approximately evaluate them using the DEIM, see Section 3.5. Since this is an approximate evaluation, we do not expect conservation of invariants, discussed in Section 7.2. However, numerical experiments in Section 7.3.3 suggest conservation of invariants when an accurate DEIM approximation is used for evaluating nonlinear terms.

To integrate (7.36) in time, the fully discrete system (7.31) is modified prior to model reduction, by dividing the mass and momentum equation with $\sqrt{\mathbf{r}}^{n+1}$. Note that since the new form is identical to (7.31), it does not affect the conserved quantities. Subsequently, a basis for $\sqrt{\mathbf{r}}$, denoted by $V_{\sqrt{\mathbf{r}}}$, is constructed. The nonlinear terms are evaluated exactly using the quadratic expansion or approximated using the DEIM.

7.3 Numerical Experiments

7.3.1 Vortex Merging

Consider the 2-dimensional incompressible Euler equation (7.14) on a square domain $\Omega = [0, 2\pi]^2$, with periodic boundary conditions. Spatial derivatives are discretized using a Fourier spectral method. To capture the fine details characterizing the solution,

256×256 modes is used. We consider the evolution of three vortices, with the initial structure given by

$$\omega = \omega_0 + \sum_{i=1}^{3} \alpha_i e^{-\frac{(x-x_i)^2 + (y-y_i)^2}{\beta^2}}. \tag{7.43}$$

Here, $\omega = \nabla \times u$ is the vorticity, (x, y) represents the spatial coordinates, (x_i, y_i) is the center of the ith vortex, α_i its maximum amplitude, and β controls the effective radius of the vortex. In this example, the center of three vortices are

$$(x_1, y_1) = (0.75\pi, \pi), (x_2, y_2) = (1.25\pi, \pi), (x_3, y_3) = (1.25\pi, 1.5\pi), \tag{7.44}$$

close to the center of the domain. Two of the vortices have a positive spin with $\alpha_1 = \alpha_2 = \pi$ and the third rotates in the opposite direction with $\alpha_3 = -0.5\pi$. The effective radius of all the vortices is set to $\beta = 1/\pi$. This arrangement of vortices is an interesting initial condition to study the process of vortex merging. This phenomenon is often a result of fast-moving dipoles of vortices with the same spin facing another vortex [32] of opposite spin. The merging process transfers the vorticity from the initial configuration into long, narrow, and spiral-shaped strips of intense vorticity [62]. The formation of such thin vorticity filaments in the fluid may pose numerical challenges, due to aliasing.

In the context of MOR, conservation of energy and stability is crucial to capturing fine structures. With the absence of natural dissipation, straight forward application of MOR techniques for the Euler equation is often unstable.

To define the initial conditions in terms of the velocity components u and the pressure p, we define a stream-function Ψ, the solution to the equation

$$-\Delta\Psi = \omega. \tag{7.45}$$

The initial velocity is then given by $\nabla \times \Psi$. To solve the stream-function problem (7.45), we require $\int_\Omega \omega \, dx = 0$. It is easily verified that this requirement implies $\omega_0 = 0.038$. The pressure is recovered by solving the related Poisson pressure equation

$$\Delta p = -\nabla \cdot S_u(u),$$

obtained by applying the divergence operator to (7.14) and using the incompressibility condition. The implicit midpoint scheme, to mimic the time integration scheme presented in (7.31), is used to integrate in time. The merging phenomenon is simulated for a total of 18 time units using a temporal step $\Delta t = 0.004$.

Figure 7.1a illustrates the evolution of the kinetic energy for the advective, divergence, and the skew-symmetric form of the high-fidelity system. It is observed that only

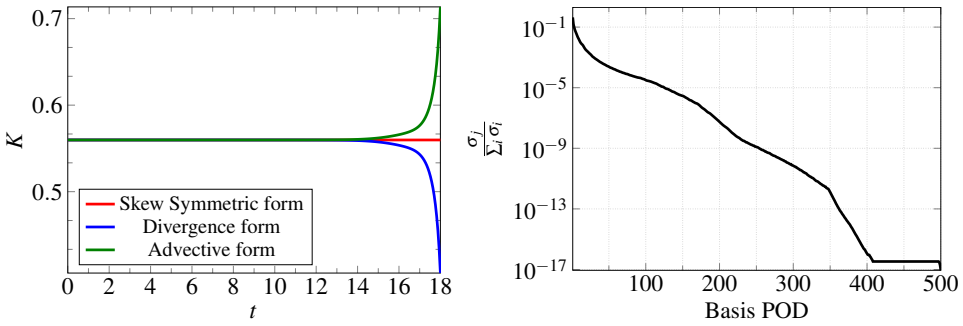


Figure 7.1 – (a) The kinetic energy K for the advective, divergence and the skew-symmetric formulations. (b) The decay of the singular values for the vortex merging.

the skew-symmetric form preserves the kinetic energy, confirming the discussion in Section 7.1.2.

A total of 5000 temporal snapshots is used to construct a reduced basis, following Algorithm 3.2. The decay of the singular values, used as an indication of the reducibility of the problem, is presented in Figure 7.1b. The first 35 POD modes corresponds to over 99% of the modes of the high fidelity solution. This suggests that an accurate reduced system can be constructed using a small number of basis vectors. To illustrate the effectiveness of the method, smaller bases are also considered.

For a qualitative analysis, in Figure 7.2, four solutions at different times are shown for the high fidelity system and the reduced system with $k = 17$ and $k = 35$ modes. The overall dynamics of the problem, and in particular the formation and development of vorticity filaments, are correctly represented, even with a moderate number of basis vectors. Although small details are not captured by the reduced system with a small number of basis vectors, the position and the spreading of the vortices are comparable.

Figure 7.3a shows the L^2 error between the high-fidelity solution and the reduced solution. The error decreases, consistently, as the number of basis vectors increases. Furthermore, the accuracy is maintained over the period of time integration.

The conservation of the kinetic energy is presented in Figure 7.3b. Even for a small number of basis vectors, where the solution is not well approximated, the kinetic energy remains constant. Furthermore, the error in the kinetic energy, due to MOR, is constant in time. This is central for the robustness of the reduced system during long time-integration.

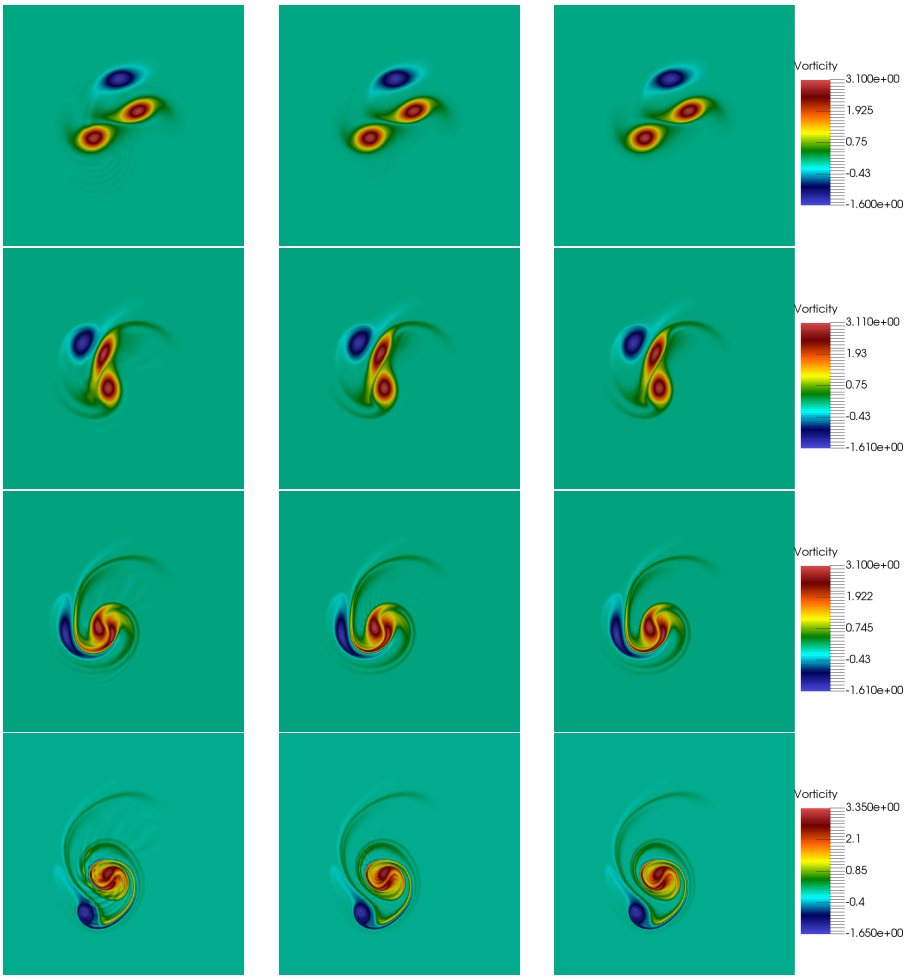
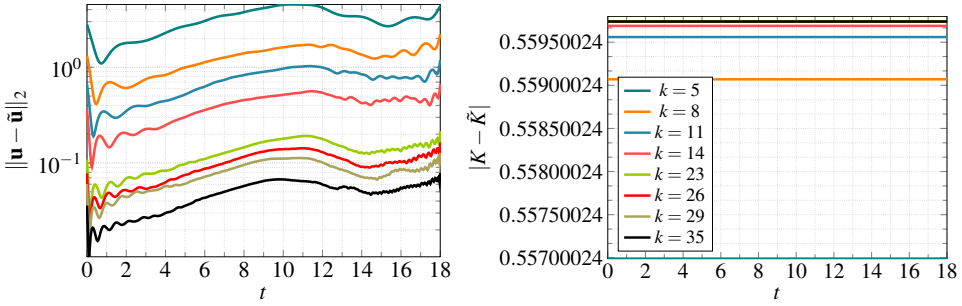


Figure 7.2 – Snapshots of the high-fidelity system and the reduced system at $t = \{4, 8, 12, 18\}$. From left to right: the solution of the reduced model with $k = 17$, $k = 35$ and the high fidelity solution.

7.3.2 2D Kelvin-Helmholtz instability

Consider the 2-dimensional compressible Euler equation (7.22) in a periodic square box $[0, 1]^2$. Unlike the incompressible example in Section 7.3.1, a centered finite difference scheme of fourth order is used to discretize (7.22). The physical domain is discretized into a grid of 256×256 nodes.



(a) (b)

Figure 7.3 – (a) Evolution of L_2 error in velocity, between the high-fidelity system and the reduced system. (b) Conservation of the kinetic energy.

The initial conditions are given by

$$\begin{cases} \mathbf{r} = \begin{cases} 2, \text{ if } 0.25 < y < 0.75, \\ 1, \text{ otherwise}, \end{cases} \\ \mathbf{u}_x = a\sin(4\pi y)\left(e^{-\frac{(y-0.25)^2}{2\sigma^2}} + e^{-\frac{(y-0.75)^2}{2\sigma^2}}\right), \\ \mathbf{u}_y = \begin{cases} 0.5, \text{ if } 0.25 < y < 0.75, \\ -0.5, \text{ otherwise}, \end{cases} \\ \mathbf{p} = 2.5, \end{cases}$$

where $a = 0.1$ and $\sigma = 5\sqrt{2}\cdot 10^{-3}$. This corresponds to contacting streams of fluid with different densities. For specific choices of parameters describing the jets, fine structures and vortices emerges at the interface between the streams. Such an instability is referred to as the Kelvin-Helmholtz instability [25].

As centered schemes are often dissipation free, resolving the discontinuous initial data requires some artificial viscosity. In the high-fidelity model, the method discussed in [112] is used as an artificial viscosity. However, at the level of the reduced system, this is replaced with a low pass filter on the expansion coefficients of POD basis vectors.

The fully discrete skew-symmetric form (7.31) is used as a time marching scheme with $\Delta t = 5\cdot 10^{-4}$ over a period of 1 time unit.

Figure 7.4 illustrates that the accuracy of the method consistently improves as a higher number of POD basis modes are considered. Furthermore, the skew-symmetric form preserves the accuracy over the period of time integration. It is observed in Figure 7.6 that all features of the flow are correctly represented in the reduced system, even with a low number of basis vectors.

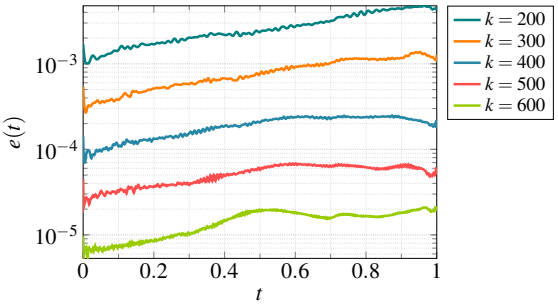


Figure 7.4 – Evolution in time of the error between the high fidelity solution of the Kelvin-Helmoltz and the reduced solution for different number of basis k. As error measure we use $e(t) = \sqrt{\|\mathbf{r} - \mathbf{r}^r\|^2 + \|\mathbf{u_x r} - \mathbf{u_x r}^r\|^2 + \|\mathbf{u_y r} - \mathbf{u_y r}^r\|^2 + \|\mathbf{p} - \mathbf{p}^r\|^2}$.

Conservation of mass, momentum and energy is presented in Figure 7.5. The accuracy of the method in approximating these invariants improves as the size of the basis is increased. Furthermore, Figure 7.5c shows how the kinetic energy associated with the reduced system mimic the kinetic energy of the high-fidelity system. This helps to ensure the correct evolution of kinetic energy, and thus, the internal energy.

7.3.3 1D Shock problem

In this section we study the 1-dimensional compressible Euler problem, (7.22) without viscous terms, with a steady state discontinuous solution. This is in preparation for Section 7.3.4, where development and propagation of shock waves is discussed. Here we asses how the skew-symmetric form of (7.22) can recover moving discontinuities at the level of the reduced system. Consider a periodic boundary conditions on $\Omega = [0, 1]$ with the initial condition

$$\begin{cases} \mathbf{r} = 0.5 + 0.2\cos(2\pi x), \\ \mathbf{u} = 1.5, \\ \mathbf{p} = 0.5 + 0.2\sin(2\pi x). \end{cases}$$

The domain is discretized into $N = 2000$ nodes and a centered finite differences scheme is used to assemble the discrete Euler equation in skew-symmetric form, as discussed in Section 7.1.3.

The fully discrete skew-symmetric form (7.31) is used for time integration over a time interval $[0, 0.3]$. To resolve the discontinuous solution we use an artificial viscosity with $\tau = \mu \partial u / \partial x$, where $\mu = 0.5 \cdot 10^{-4}$.

Figure 7.7 shows the evolution of conserved quantities for the high-fidelity and reduced system. Here, the high-fidelity model is also considered in the divergence and advective form in addition to the skew-symmetric form. It is clear that when the reduced

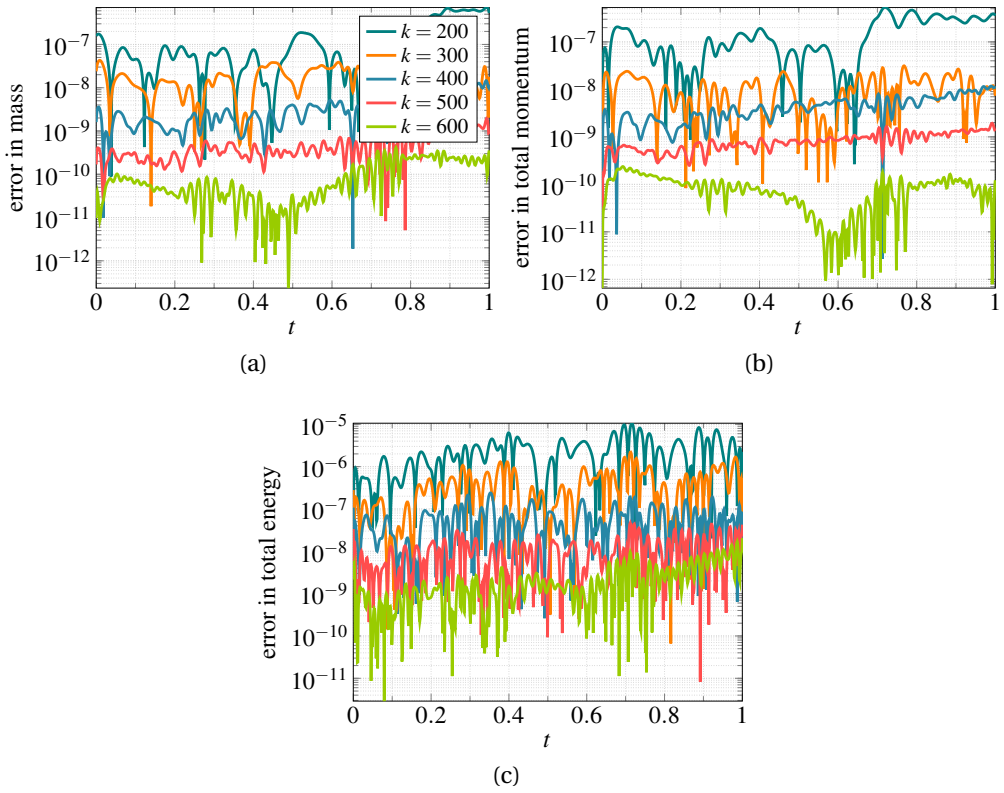


Figure 7.5 – Difference between the high fidelity solution of the Kelvin-Helmholtz problem and the reduced solution of the mass (a), the momentum (b), and the total energy (c).

systems is not in skew-symmetric form, it violates conservation of mass, momentum, and energy. Even while the high-fidelity systems in divergence and advective forms are stable, the constructed reduced system is unstable, independent on the number of basis vectors. On the other hand, the skew-symmetric form yields a stable and conservative reduce system. Note that the loss in the energy associated with the skew-symmetric form, illustrated in Figures 7.7b, 7.7d and 7.7f, is due to the application of an artificial viscosity.

Figure 7.8 shows the total error, when the reduced system captures a discontinuous solution at $t = 0.16$. It is observed that the formation of a discontinuity affects the accuracy of the method. This is expected as a sharp gradient is approximated by a relatively few POD modes. However, the method remains robust and stable during the period of time integration.

In Figure 7.9 we compare the numerical artifacts of different formulations of the Euler equation. The advective formulation is not presented since it does not yield a stable reduced system. It is observed that the reduced system based on the skew-symmetric

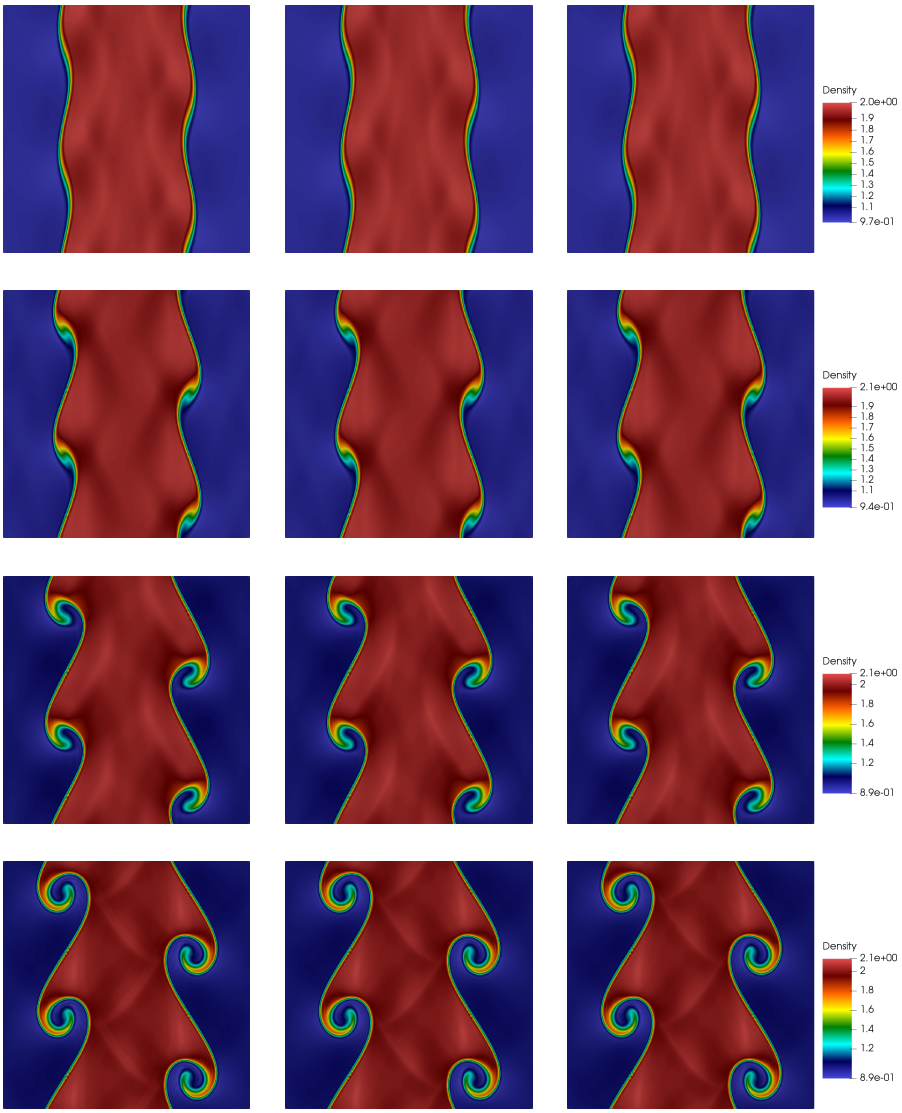


Figure 7.6 – Solutions of the Kelvin-Helmholtz problem at $t = \{0.4, 0.6, 0.8, 1\}$. From left to right we show the solution of the reduced model with $k = 200$, $k = 500$ and the high fidelity solution.

formulation accurately represent the overall behavior of the high-fidelity solution. On the other hand, a Gibbs-type error [103] appears near sharp gradients, for the reduced system based on the divergence form of the Euler equation. The well-representation of the skew-symmetric form is due the low aliasing error property.

As discussed in Section 7.2, the DEIM approximation needed for an efficient evaluation of the nonlinear components of (7.22), can affect the conservation properties of the skew-symmetric form. Figure 7.10 shows the decay of the singular values of the nonlinear snapshots. The decay of these snapshots is significantly slower than the temporal snapshots of (7.22). This indicates that to maintain the accuracy of the reduced

system, the DEIM basis should be chosen richer than the POD basis. Figure 7.11a and Figure 7.11b present the error and the conservation of total energy when the DEIM is used to approximate the nonlinear term. The conservation of energy is recovered once DEIM approximates the nonlinear terms with enough accuracy. In this numerical experiment, evaluation of the nonlinear terms in (7.22) using the DEIM is ten times faster than the high-fidelity evaluation.

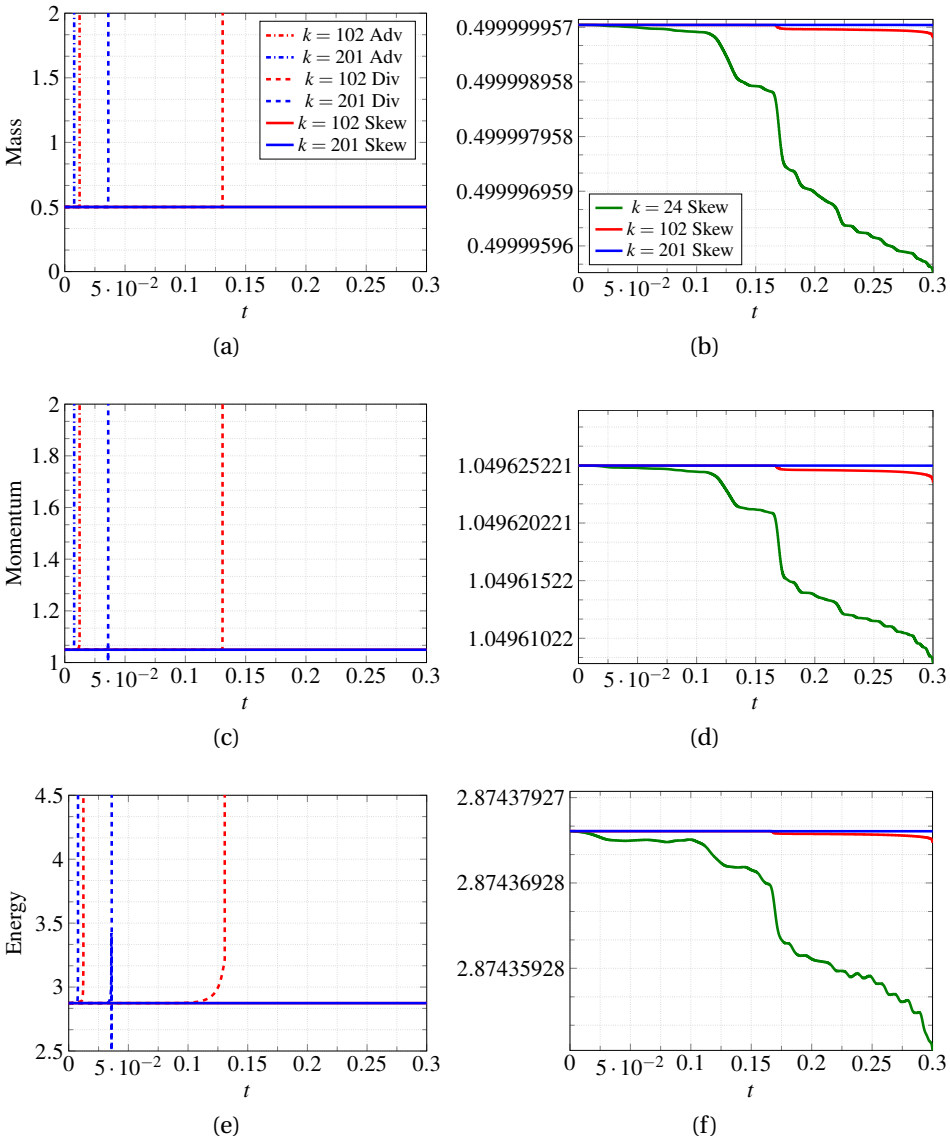


Figure 7.7 – (left) Evolution of the three conserved quantities for the reduced solution of the compressible Euler equation (mass, total momentum and total energy). The divergent, advective and skew-symmetric formulations have been considered and $k = 102, 204$ basis are used in the reduced model. (right) Evolution of the conserved quantity for a stable reduced model using the skew-symmetric formulation.

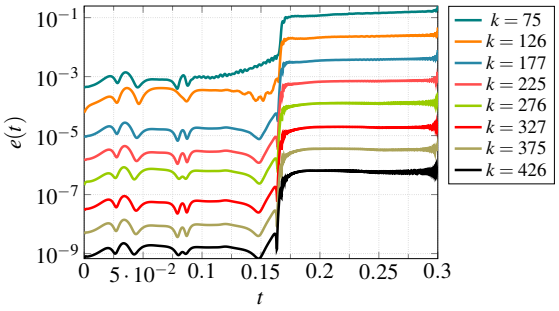


Figure 7.8 – Evolution in time of the error between the high fidelity solution of the 1D compressible Euler and the reduced solution for different number of basis k. As error measure we consider $e(t) = \sqrt{\|\mathbf{r} - \mathbf{r}^r\|^2 + \|\mathbf{ur} - \mathbf{ur}^r\|^2 + \|\mathbf{p} - \mathbf{p}^r\|^2}$.

7.3.4 Continuous Variable Resonance Combustor

CVRC is a model rocket combustor, designed and operated at Purdue University (Indiana, U.S.) to investigate combustion instabilities [113]. This setup is called the Continuously Variable Resonance Combustor (CVRC) because the length of the oxidizer injector can be varied continuously, allowing for a detailed investigation of the coupling between acoustics and combustion in the chamber [45]. The 2D/3D high-fidelity simulations of CVRC are expensive. Thus to get a fast analysis tool, a quasi-1D model has been proposed by Smith et al. [99] and further developed by Frezzotti et al. [42, 41, 43].

The CVRC consists of three parts: oxidizer post, combustion chamber and exit nozzle, as shown in Fig. 7.12. The oxidizer is injected from the left end of the oxidizer post and meets the fuel, injected through an annular ring around the oxidizer injector, at the back-step. The combustion happens in a region around the back-step. The combustion products flow through the chamber and exit the system from the nozzle. Both the injector and the nozzle are operated at choked condition during the experiment. The length of the oxidizer post L_{op} of the CVRC can be varied continuously, leading to different dynamics. Here, we focus on the case with $L_{op} = 14.0$ cm, in which the combustion is unstable.

The geometry parameters of the quasi-1D CVRC with a oxidizer post length $L_{op} = 14.0$ cm are shown in Table 7.1. The back-step and the converging part of the nozzle are sinusoidally contoured to avoid a discontinuity of the radius that will invalidate the quasi-1D governing equations presented in the next subsection.

The fuel is pure gaseous methane. The oxidizer is a mixture of 42% oxygen and 58% water (per unit mass), and is injected in the oxidizer post at a temperature $T_{ox} = 1030$K so that both water and oxygen are in the gaseous phase. The operating conditions are listed in Table 7.2.

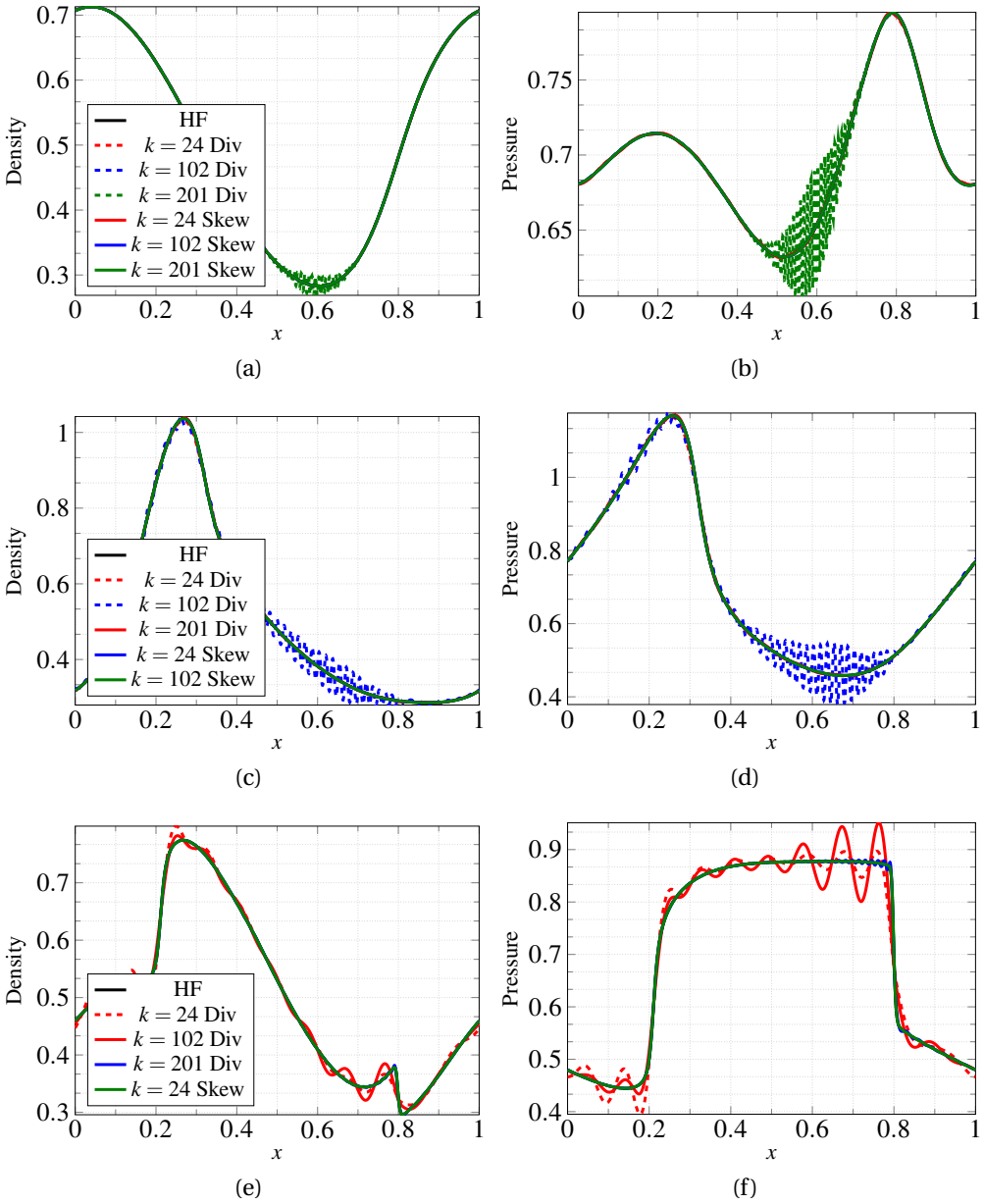


Figure 7.9 – Qualitative comparison between different formulations for the reduced model in terms of density (left) and pressure (right) at $t = 0.1, 0.3$ and 1s. Results for the advective formulation are not showed here because the related reduced solutions are unstable after a few time steps.

For the combustion, we consider the one-step reaction model

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O.$$

We assume that the fuel reacts instantaneously to form products, allowing us to neglect

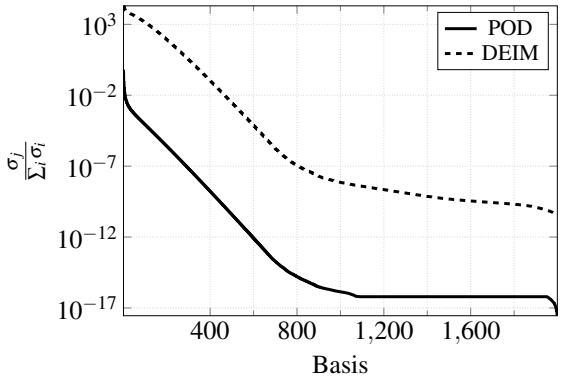


Figure 7.10 – Decay of the singular values of the snapshot matrix related to POD and DEIM algorithms for the 1D compressible Euler problem.

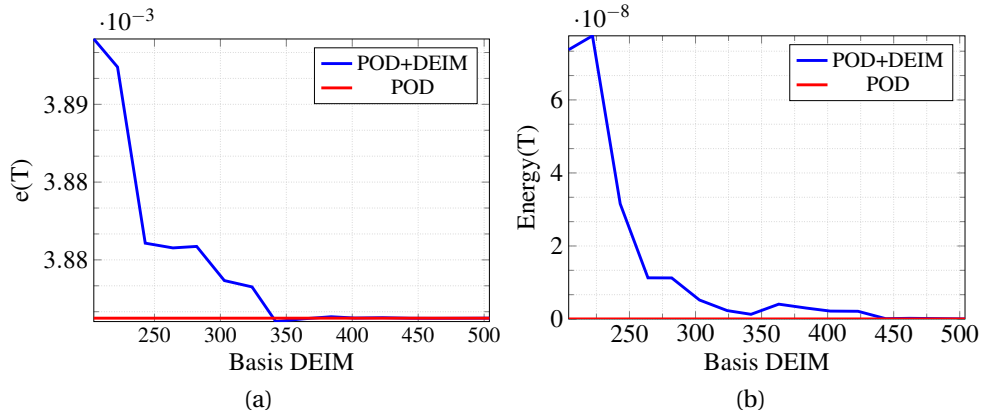


Figure 7.11 – Comparison between standard POD and POD with DEIM treatment of the nonlinear term in terms of the error (a) and the total energy (b).

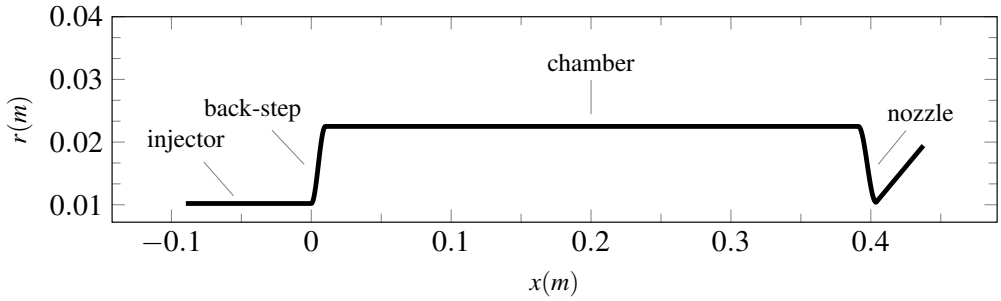


Figure 7.12 – Geometry of quasi-1D CVRC model.

intermediate species and finite reaction rates. As the equivalence ratio is less than one, there is oxidizer left after the combustion. Therefore, only two species need to be considered: oxidizer and combustion products.

The governing equations that describe the conservation of mass, momentum, and

Table 7.1 – Geometry parameters of the quasi-1D CVRC with an oxidizer post length $L_{op} = 14$ cm.

Section	Oxidizer post		Chamber	Nozzle	
	injector	back-step		converging part	diverging part
Length (cm)	12.99	1.01	38.1	1.27	3.4
Radius (cm)	1.02	$1.02 \sim 2.25$	2.25	$2.25 \sim 1.04$	$1.04 \sim 1.95$

Table 7.2 – CVRC operating conditions.

Parameter	Unit	Value
Fuel mass flow rate, $\dot{m}_f$	kg/s	0.027
Fuel temperature, T_f	K	300
Oxidizer mass flow rate, $\dot{m}_{ox}$	kg/s	0.32
Oxidizer temperature, T_{ox}	K	1030
O_2 mass fraction in oxidizer, Y_{O_2}	–	42.4%
H_2O mass fraction in oxidizer, Y_{H_2O}	–	57.6%
Mean chamber pressure	MPa	1.34
Equivalence ratio, E_r	–	0.8

energy of the quasi-1D CVRC flow, are the quasi-1D unsteady Euler equations for multiple species, expressed in conservative form as

$$\frac{\partial}{\partial t} v + \frac{\partial}{\partial x} F_v = s_A + s_f + s_q. \tag{7.46}$$

The conserved variable vector v and the convective flux vector F are

$$v = \begin{pmatrix} \rho A \\ \rho u A \\ \rho E A \\ \rho Y_{ox} A \end{pmatrix}, F = \begin{pmatrix} \rho u A \\ \left(\rho u^2 + p\right) A \\ \left(\rho E + p\right) u A \\ \rho u Y_{ox} A \end{pmatrix}, \tag{7.47}$$

where ρ is the density, u is the velocity, p is the pressure, E is the total energy, Y_{ox} is the mass fraction of oxidizer, and $A = A(x)$ is the cross sectional area of the duct. The pressure p can be computed using the conserved variables as

$$E = \frac{p}{\rho(\gamma - 1)} + \frac{u^2}{2} - C_p T_{ref}, \tag{7.48}$$

where T_{ref} is the reference temperature and is set as 298.15 K. The temperature T is recovered from the equation of state $p = \rho RT$. The gas properties C_p, R and γ are

computed as $C_p = \sum C_{pi} Y_i$, $R = \sum R_i Y_i$ and $\gamma = C_p/(C_p - R)$, respectively.

The source terms are

$$s_A = \begin{pmatrix} 0 \\ p\dfrac{dA}{dx} \\ 0 \\ 0 \end{pmatrix}, s_f = \begin{pmatrix} \dot{\omega}_f \\ \dot{\omega}_f u \\ \dot{\omega}_f \left(h_0^f + \Delta h_0^{rel} \right) \\ \dot{\omega}_{ox} \end{pmatrix}, s_q = \begin{pmatrix} 0 \\ 0 \\ q' \\ 0 \end{pmatrix}, \tag{7.49}$$

where $\dot{\omega}_f$ is the depletion rate of the fuel, $\dot{\omega}_{ox}$ is the depletion rate of the oxidizer, h_0^f is the total enthalpy of the fuel, Δh_0^{rel} is the heat of reaction per unit mass of fuel and q' is the unsteady heat release term. s_A accounts for area variations, s_f and s_q are related to the combustion. s_f represents the addition of the fuel and its combustion with the oxidizer, which in turn results in the creation of the combustion products. The depletion rate of the fuel is

$$\dot{\omega}_f = \frac{k_f \dot{m}_f Y_{ox} \left(1 + sin\xi\right)}{l_f - l_s}, \tag{7.50}$$

where

$$\xi = -\frac{\pi}{2} + 2\pi \frac{x - l_s}{l_f - l_s}, \quad \forall \ l_s < x < l_f. \tag{7.51}$$

The setting of the fuel injection restricts the combustion to the region $l_s < x < l_f$. The reaction constant k_f is selected to insure that the fuel is consumed within the specified combustion zone. The depletion rate of the oxidizer is computed by

$$\dot{\omega}_{ox} = C_{o/f} \dot{\omega}_f, \tag{7.52}$$

where $C_{o/f}$ is the oxidizer-to-fuel ratio.

The unsteady heat release term q', also called the combustion response function, models the coupling between acoustics and combustion. Here, we use the combustion response function designed by Frezzotti et al. [41, 43], which is a function of the velocity, sampled at specific abscissa $\hat{x}$ that is almost coincident with the antinode of the first longitudinal modal shape with a time lag t_0, i.e.,

$$q'\left(x,t\right) = \alpha g\left(x\right) A\left(x\right) \left[u\left(\hat{x}, t - t_0\right) - \bar{u}\left(\hat{x}\right)\right]. \tag{7.53}$$

Here $\bar{u}$ is the time averaged velocity, estimated with the steady-state quasi-1D model

assuming $q' = 0$, and $g(x)$ is a Gaussian distribution

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{7.54}$$

where μ is the mean and σ is the standard deviation. The amount of heat release due to velocity oscillations is controlled by the parameter α, in (7.53).

The boundary conditions for the quasi-1D CVRC flow include the fixed mass flow rate and the stagnation temperature at the head-end of the oxidizer injector, and the supersonic outflow at the exit of the nozzle.

Prior to the unsteady simulation, the quasi-1D CVRC needs to be excited, which is achieved by adding a perturbation to the steady-state solution. The perturbation is added by forcing the mass flow rate with a multi-sine signal

$$\dot{m}_{ox}(t) = \dot{m}_{ox,0}\left[1 + \delta \sum_{k=1}^{K} sin(2\pi k \Delta f t)\right], \tag{7.55}$$

where $\dot{m}_{ox,0}$ is the oxidizer mass flow rate in Table 7.2, Δf is the frequency resolution and K is the number of frequencies. In this paper, $\Delta f = 50$ Hz and $K = 140$, resulting in a minimal frequency of 50 Hz and a maximal frequency of 7000 Hz. δ is required to be small to control the amplitude of the perturbation and is set as 0.1%.

The procedure of the unsteady simulation of the quasi-1D CVRC flow includes three steps:

1. Compute the steady-state solution by setting $\dot{m}_{ox} = \dot{m}_{ox,0}$ and $q' = 0$.
2. Excite the system by adding a perturbation to the oxidizer mass flow rate according to (7.53) and setting $q' = 0$.
3. Perform the unsteady simulation by turning on the combustion response function q' in (7.49) and turning off the oxidizer mass flow rate perturbation by setting $\dot{m}_{ox} = \dot{m}_{ox,0}$.

Introduction of an artificial viscosity is essential for a robust and long time-integration of (7.49). Common discretization schemes for (7.49) are often dissipative, e.g., the Lax-Friedrich scheme used in [108]. Since the skew-symmetric discretization is non-dissipative, we modify (7.49) as

$$\frac{\partial}{\partial t} v + \frac{\partial}{\partial x} F = s_A + s_f + s_q + d, \quad d = (0, \frac{\partial}{\partial x}\tau, 0, 0)^T, \tag{7.56}$$

with $\tau = \mu \partial(uA)/\partial x$, and $\mu = 6 \times 10^{-5}$. This type of artificial viscosity is chosen for its

simplicity. This, however, can be replaced with a more moderate and sophisticated method.

Note that the right hand side in (7.56) suggests that, in general, mass, momentum, and energy is not conserved. Furthermore, the complex coupling of the variables in (7.49) and the non-constant adiabatic gas index prohibit the application of complex and implicit time integration schemes. Therefore, a quasi-skew-symmetric form, introduced in (7.12), is used for (7.49). It is straight-forward to check [98], for $t, s \in \mathbb{R}^N$

$$\frac{1}{2}\delta_x(st)_j + \frac{1}{2}s_j\delta_x(t)_j + \frac{1}{2}t_j\delta_x(s)_j = \frac{1}{4}\delta_x^+(s_j + s_{j-1})(t_j + t_{j-1}). \tag{7.57}$$

where $\delta_x(v)_j = (v_{j+1} - v_{j-1})/\Delta x$ is centered finite difference approximation of the space derivative and $\delta_x^+(v_j) = (v_{j+1} - v_j)/\Delta x$, for some $v \in \mathbb{R}^N$. Therefore,

$$F^{\Delta}_{i+1/2}(s_jt_j, s_{j+1}t_{j+1}) = (s_j + s_{j-1})(t_j + t_{j-1}), \tag{7.58}$$

can be interpreted as an approximation of a quadratic flux function at the boundary of two adjacent finite volume cells. A better approximation of the flux in (7.58) corresponds to a higher order skew-symmetric form for a quadratic variable st in (7.57). We discretize the real line into N uniform cells of size Δx. A quasi-skew-symmetric form for (7.56) now takes the form

$$\begin{aligned}\frac{d}{dt}q_j^i + \delta^+F^{\Delta}_{i+1/2}(q_j^ir_j^i, q_{j+1}^ir_{j+1}^i) - \delta^+F_d^{\Delta}(d_j^i, d_{j+1}^i) + \delta^+F_p^{\Delta}(p_j, p_{j+1}) \\ = \int_{c_j} s_A + s_f + s_q\, dx.\end{aligned} \tag{7.59}$$

for $j = 1, \ldots, N$. Here, c_j is the jth cell, $q_j^i = \int_{c_j} v^i\, dx$ is the cell average of the ith component of v, F_p^{Δ} is the flux approximation of the pressure term, F_d^{Δ} is the flux approximation for the viscous term and $r = (u, u, u, u)^T$.

The three-stage Runge-Kutta (SSP RK3) [59] is used to integrate (7.56) in time. The pressure profile for the steady state, with $q' = 0$, and the pressure oscillatory mode in the unsteady phase is presented in Figures 7.13a and 7.13b, respectively.

The discontinuities that appear in the solution of (7.56) suggests that a relatively large basis is required to resolve fine structures in the solution. Here, a POD basis is generated with $k = 200$, $k = 300$ and $k = 400$ number of basis vectors. To avoid basis changes in the reduced system, only one POD basis is considered for ρ, ρu ρE and ρY_{ox}. The explicit SSP RK3 is then used to integrated the reduced system in time, for the unsteady system. The source terms are evaluated in the high-fidelity space and projected onto the reduced space. However, in principle, the DEIM can be applied to accelerate the evaluation this component.

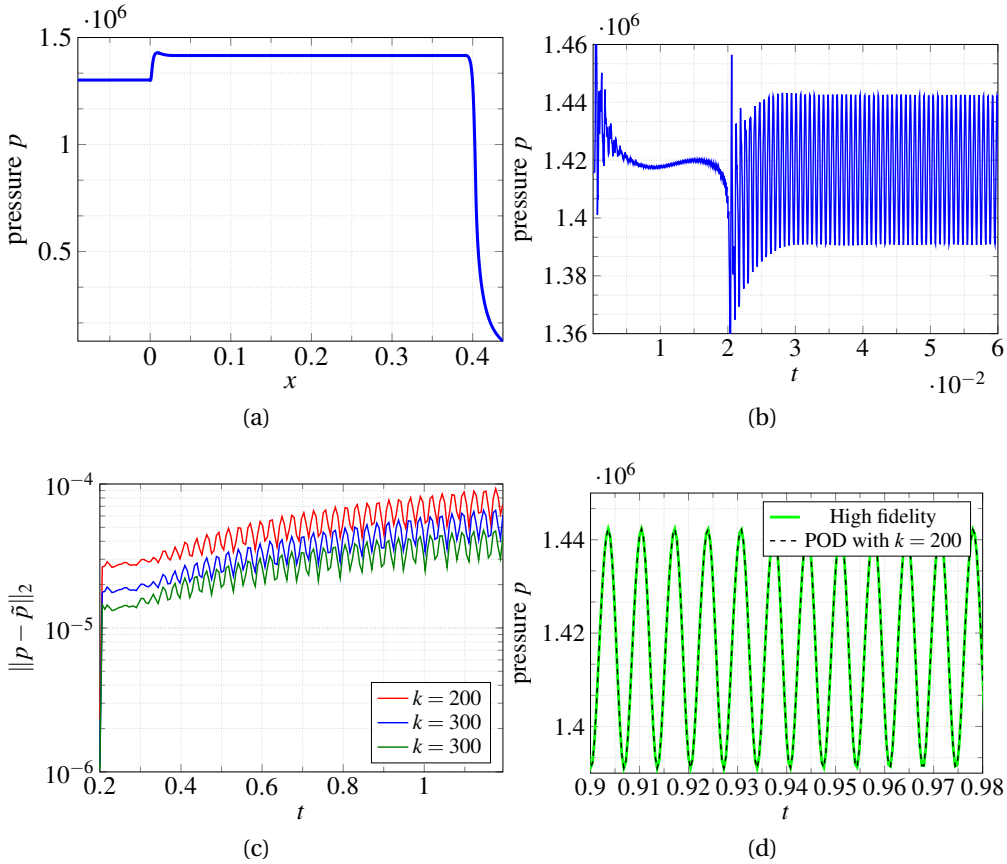


Figure 7.13 – (a) Pressure profile of the steady state. (b) Oscillatory mode of pressure located at $x = 0.36$ for the unsteady flow. (c) Relative error between the high-fidelity and approximated pressure. (d) Approximation of the oscillations.

Figure 7.13c shows the approximation error of the pressure, due to MOR. It is observed that the approximation is consistently improved as the number of basis vectors increases. Furthermore, the approximate solution maintains high accuracy over a relatively long time-integration. The oscillations of pressure is demonstrated in Figure 7.13d. The overall behaviour of pressure is well approximated by the reduced system. Similar results are obtained for a POD basis with higher number of modes.

We note that the discrete form of (7.56) is not in the full skew-symmetric form. Nonetheless, the quasi-skew-symmetric discretization offers remarkable stability preservation.

7.4 Conclusions

Conservation of nonlinear invariants are not, in general, guaranteed with conventional model reduction techniques. The violation of such invariants often result in a qualitatively wrong or unstable reduced system, even when the high-fidelity system is stable.

This is particularly important for fluid flow, where conservation of the energy, as a nonlinear invariant of the system, is crucial for a correct numerical evaluation.

In this paper, we discuss that conservative properties of the skew-symmetric form for fluid flow can naturally be extended to the reduced system. Conventional MOR techniques preserves the skew-symmetry of differential operator which result in the conservation of quadratic invariants at the level of the reduced system. Furthermore, the reduced system also contains quadratic invariants with respect to the reduced variables that approximates the invariants of the high-fidelity system. This results in the construction of a physically meaningful reduced system, rather than a mere couple systems of differential equations.

Numerical experiments for the incompressible and compressible Euler equation confirms conservation of mass, momentum and energy for the reduced model with the skew-symmetric discretization. In contrast, when a non-skew-symmetric form, e.g. divergence form or advective form, is considered, MOR does not necessarily yield a stable reduced system. On the other hand the skew-symmetric form consistently yields a robust reduced system over long time-integration, even when the reduced space does not represent the high-fidelity solution accurately.

Finally, a MOR of a quasi-skew-symmetric form for the CVRC model is presented. Although this model is not in a full skew-symmetric form and an explicit Runge-Kutta method used for time-integration, we still recover a reduced model with excellent stability properties.

8 Conclusions

During the past decades, the need to solve complex, multi-physics, and multi-scale applications have become central in science, engineering and across many industrial domains. The numerical evaluation of such models using classical approaches, however, is often prohibitive due to limitations in computational capacities. In such situations, model order reduction is playing an increasingly important role in advancements in scientific computing and high-performance computing by reducing the intrinsic computational complexity of many modern models.

Despite the success of model order reduction for elliptic and parabolic PDEs, model reduction for hyperbolic systems remains a challenge. Symmetries, invariants, and conservation laws are a fundamental feature of such models, and are often destroyed during model order reduction. The violation of such features not only result in an inaccurate model, but may also cause numerical instabilities in the reduced order model.

This thesis studies and develops model order reduction techniques that conserve certain invariants and symmetries of hyperbolic systems of PDEs. Conserving such structures not only result in a physically meaningful reduced model, but provides robust long time behaviour and a stable reduced model.

To achieve this goal, we study model order reduction from a geometric point of view. The crucial role of time is highlighted for the construction of symmetry-preserving model order reduction. We furthermore investigate why conventional model order reduction techniques often break the symmetries of hyperbolic problems.

Hamiltonian systems, as a special case of highly symmetric PDEs, are intensively studied in this thesis. We discuss how the symplectic structure, the symmetry of Hamiltonian systems, can be conserved in model order reduction. A greedy approach for construction of a reduced basis is presented. And we discuss how a symplectic Galerkin projection constructs a reduced Hamiltonian system that carries the symmetries of

the original Hamiltonian system. The reduced Hamiltonian, as an approximation to the original Hamiltonian, is a conserved quantity for the reduced system. Hence, the loss in the Hamiltonian due to model order reduction remains constant and can be controlled.

To adapt the symplectic model reduction to an unstructured numerical discretization, the method is coupled with a weighted norm. A reduced system is constructed by orthogonally projecting a generalized Hamiltonian system onto the reduced space, with respect to a weighted inner product. The reduced system, however, carries the Hamiltonian structure and also the symplectic symmetry. It is shown that the new method can be viewed as a natural extension of the symplectic model reduction, and therefore retains the structure preserving features, e.g. symplecticity and stability.

In many applications in engineering, models appear as a dissipative perturbation of Hamiltonian system. In such models, the Hamiltonian systems is no longer symplectic. In this thesis, we consider a canonical extension of dissipative Hamiltonian systems, by coupling the dissipative system with a canonical heat bath, resulting a closed and conservative system. A symplectic model reduction method can then be applied to conserve the symmetries of the extended model, and, consequently, conserve the evolution of energy and dissipation at the level of the reduced system. It is shown that the extension of the system does not pose a significant additional computational burden.

The numerical experiments in this thesis illustrate that the proposed methods consistently result in a robust reduced system with excellent stability. Conventional model reduction techniques, even when the reduced basis is chosen to yield a high accuracy, may yield an unstable or poorly performing reduced system. Numerical experiments confirm that the conservation of symmetries can significantly enhance the overall dynamics of the reduced system.

To generalize the symplectic model reduction to more complex problems, a conservative model reduction technique of fluid flow is proposed. Skew-symmetric models for fluid flow are well-known for conserving quadratic invariants of a fluid flow in a numerical evaluation. The key ingredient in these methods is the construction of a discrete skew-symmetric operator. A proper model order reduction method preserves the skew-symmetry of such differential operators. This helps to define quadratic invariants in the reduced system that approximate the quadratic invariants of the high fidelity system. Numerical experiments suggest that the skew-symmetric form consistently yields a robust reduced system over long time-integration, even when the reduced model does not represent the high-fidelity solution accurately.

What is less emphasised in this thesis is the question of reducibility of general hyperbolic problems. Transport of information, potentially throughout the entire domain,

is a distinctive feature of hyperbolic problems. This often covers patterns in the ensemble of snapshots of the system and inhibits the possibility of describing the system as a linear combination of a relatively few basis vectors. The construction of efficient reduced order models for such cases, therefore, can be a possible extension to this work.

Although conservation of symplectic symmetry and quadratic invariants is discussed intensively in this thesis, conservation of general conservation laws or invariants is to be investigated. An extension of symplectic model order reduction may be to seek the conservation of the Poisson structure, or the multi-symplectic structure, on a symplectic manifold. In addition, the conservation of integral curves over model order reduction, in order to recover a stable reduced system, remains future work.

This thesis provides a promising approach to the construction of robust, accurate, physically meaning-full reduced system for Hamiltonian systems and fluid flow. It is also extends the understanding of what can be achieved with model order reduction and reduced basis methods. Indeed, the conservation of general nonlinear invariants, e.g. for Hamiltonian systems, in a linearly transformed and approximated system is a key achievement.

Modeling is the art of approximately describing nature with understandable tools. Therefore, constructing a simplified and reduced model that resembles the symmetries and distinctive features and invariants of a complex system is another step in mathematical modelling. What is presented in this thesis highlights the potential role of structure-preserving model reduction in the future advancements of modeling.

Bibliography

[1] R. Abraham and J. E. Marsden. *Foundations of mechanics*, volume 36.

[2] B. M. Afkham and J. S. Hesthaven. Structure preserving model reduction of parametric hamiltonian systems. *SIAM Journal on Scientific Computing*, 39(6):A2616–A2644, 2017.

[3] D. Amsallem and C. Farhat. On the stability of reduced-order linearized computational fluid dynamics models based on POD and Galerkin projection: descriptor vs non-descriptor forms. In *Reduced order methods for modeling and computational reduction*, pages 215–233. Springer, Cham, 2014.

[4] A. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics, 2005.

[5] J. A. Atwell and B. B. King. Proper orthogonal decomposition for reduced basis feedback controllers for parabolic equations. *Mathematical and computer modelling*, 33(1-3):1–19, 2001.

[6] F. Ballarin, A. Manzoni, A. Quarteroni, and G. Rozza. Supremizer stabilization of POD–Galerkin approximation of parametrized steady incompressible Navier–Stokes equations. *International Journal for Numerical Methods in Engineering*, 102(5):1136–1161, 2015.

[7] M. F. Barone, I. Kalashnikova, D. J. Segalman, and H. K. Thornquist. Stable galerkin reduced order models for linearized compressible flow. *Journal of Computational Physics*, 228(6):1932–1946, 2009.

[8] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation'method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[9] C. Beattie and S. Gugercin. Structure-preserving model reduction for nonlinear port-Hamiltonian systems. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 6564–6569. IEEE, 2011.

[10] P. Benner and T. Breiten. Interpolation-based $\mathcal{H}_2$-model reduction of bilinear control systems. *SIAM Journal on Matrix Analysis and Applications*, 33(3):859–885, 2012.

[11] P. Benner, R. Byers, H. Faßbender, V. Mehrmann, and D. Watkins. Cholesky-like factorizations of skew-symmetric matrices. *Electronic Transactions on Numerical Analysis*, 11:85–93 (electronic), 2000.

[12] P. Benner and P. Goyal. *An Iterative Model Reduction Scheme forÂ Quadratic-Bilinear Descriptor Systems with an Application to Navier–Stokes Equations*, pages 1–19. Springer International Publishing, Cham, 2018.

[13] P. Benner, V. Mehrmann, and H. Xu. A new method for computing the stable invariant subspace of a real hamiltonian matrix. *Journal of computational and applied mathematics*, 86(1):17–43, 1997.

[14] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox. *Model Reduction and Approximation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

[15] N. Bhatia and G. Szegö. *Stability Theory of Dynamical Systems*. Classics in Mathematics. Springer Berlin Heidelberg, 2002.

[16] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM Journal on Mathematical Analysis*, 43(3):1457–1472, 2011.

[17] G. A. Blaisdell. *Numerical Simulations of Compressible Homogeneous Turbulence*. PhD thesis, Stanford University, 1991.

[18] S. Blanes and F. Casas. *A concise introduction to geometric numerical integration*, volume 23. CRC Press, 2016.

[19] A. Buffa, Y. Maday, A. T. Patera, C. Prud'homme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(3):595–603, 2012.

[20] A. Bunse-Gerstner. Matrix factorizations for symplectic qr-like methods. *Linear Algebra and its Applications*, 83:49–77, 1986.

[21] K. Carlberg, Y. Choi, and S. Sargsyan. Conservative model reduction for finite-volume models, 2017.

[22] K. Carlberg, Y. Choi, and S. Sargsyan. Conservative model reduction for finite-volume models. *Journal of Computational Physics*, 371:280–314, 2018.

[23] K. Carlberg, R. Tuminaro, and P. Boggs. Efficient structure-preserving model reduction for nonlinear mechanical systems with application to structural dynamics. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, page 1969.

[24] K. Carlberg, R. Tuminaro, and P. Boggs. Preserving lagrangian structure in nonlinear model reduction with application to structural dynamics. *SIAM Journal on Scientific Computing*, 37(2):B153–B184, 2015.

[25] S. Chandrasekhar. Hydrodynamic and hydromagnetic stability. 2013.

[26] S. Chaturantabut, C. Beattie, and S. Gugercin. Structure-preserving model reduction for nonlinear port-hamiltonian systems. *SIAM Journal on Scientific Computing*, 38(5):B837–B865, 2016.

[27] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[28] C. Corduneanu. *Integral equations and applications*, volume 148. Cambridge University Press Cambridge, 1991.

[29] N. N. Cuong, K. Veroy, and A. T. Patera. Certified real-time solution of parametrized partial differential equations. In *Handbook of Materials Modeling*, pages 1529–1564. Springer, 2005.

[30] S. Deparis and G. Rozza. Reduced basis method for multi-parameter-dependent steady Navier–Stokes equations: Applications to natural convection in a cavity. *Journal of Computational Physics*, 228(12):4359 – 4378, 2009.

[31] O. Desjardins, G. Blanquart, G. Balarac, and H. Pitsch. High order conservative finite difference scheme for variable density low mach number turbulent flows. *Journal of Computational Physics*, 227(15):7125–7159, 2008.

[32] D. G. Dritschel and N. J. Zabusky. On the nature of vortex interactions and models in unforced nearly inviscid two dimensional turbulence. *Physics of Fluids*, 8(5):1252–1256, 1996.

[33] L. Edsberg. *Introduction to computation and modeling for differential equations*. John Wiley & Sons, 2015.

[34] E. Faou. *Geometric Numerical Integration and Schrödinger Equations*. Zurich lectures in advanced mathematics. European Mathematical Society, 2012.

[35] S. C. Farantos. *Nonlinear Hamiltonian Mechanics Applied to Molecular Dynamics: Theory and Computational Methods for Understanding Molecular Spectroscopy and Chemical Reactions*. Springer, 2014.

[36] C. Farhat, T. Chapman, and P. Avery. Structure-preserving, stability, and accuracy properties of the energy-conserving sampling and weighting method for the hyper reduction of nonlinear finite element dynamic models. *International Journal for Numerical Methods in Engineering*, 102(5):1077–1110, 2015.

[37] J. Fehr, D. Grunert, A. Bhatt, and B. Haasdonk. A sensitivity study of error estimation in elastic multibody systems. In A. Kugi, editor, *Proceedings 9th Vienna International Conference on Mathematical Modelling, MATHMOD 2018*, 2018.

[38] F. Feppon and P. F. J. Lermusiaux. A geometric approach to dynamical model order reduction. *SIAM Journal on Matrix Analysis and Applications*, 39(1):510–538, 2018.

[39] A. Figotin and J. H. Schenker. Spectral theory of time dispersive and dissipative systems. *Journal of statistical physics*, 118(1-2):199–263, 2005.

[40] A. Figotin and J. H. Schenker. Hamiltonian structure for dispersive and dissipative dynamical systems. *Journal of Statistical Physics*, 128(4):969–1056, 2007.

[41] M. L. Frezzotti, S. D'Alessandro, B. Favini, and F. Nasuti. Numerical issues in modeling combustion instability by quasi-1D euler equations. *International Journal of Spray and Combustion Dynamics*, 9(4):349–366, 2017.

[42] M. L. Frezzotti, F. Nasuti, C. Huang, C. Merkle, and W. E. Anderson. Determination of heat release response function from 2D hybrid RANS-LES data for the CVRC combustor. In *51st AIAA/SAE/ASEE Joint Propulsion Conference*, page 3841, 2015.

[43] M. L. Frezzotti, F. Nasuti, C. Huang, C. L. Merkle, and W. E. Anderson. Quasi-1D modeling of heat release for the study of longitudinal combustion instability. *Aerospace Science and Technology*, 75:261–270, 2018.

[44] A. Friedman. *Foundations of Modern Analysis*. Dover Books on Mathematics Series. Dover, 1970.

[45] R. Garby. *Simulations of flame stabilization and stability in high-pressure propulsion systems*. PhD thesis, INPT, 2013.

[46] S. Gugercin, R. V. Polyuga, C. Beattie, and A. Van Der Schaft. Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. *Automatica*, 48(9):1963–1974, 2012.

[47] B. Haasdonk. Convergence rates of the pod–greedy method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(3):859–873, 2013.

[48] B. Haasdonk. Reduced basis methods for parametrized PDEs – a tutorial introduction for stationary and instationary problems. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, pages 65–136. SIAM, Philadelphia, 2017.

[49] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(2):277–302, 2008.

[50] B. Haasdonk and M. Ohlberger. Efficient reduced models and a-posteriori error estimation for parametrized dynamical systems by offline/online decomposition. *Mathematical and Computer Modelling of Dynamical Systems*, 17(2):145–161, 2011.

[51] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, volume 31. Springer, 2006.

[52] J. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. SpringerBriefs in Mathematics. Springer International Publishing, 2015.

[53] A. E. Honein. *Numerical aspects of compressible turbulence simulations*. 2005.

[54] A. E. Honein and P. Moin. Higher entropy conservation and numerical stability of compressible turbulence simulations. *Journal of Computational Physics*, 201(2):531–545, 2004.

[55] K. Ito and S. Ravindran. A reduced basis method for control problems governed by pdes. In *Control and estimation of distributed parameter systems*, pages 153–168. Springer, 1998.

[56] K. Ito and S. Ravindran. A reduced-order method for simulation and control of fluid flows. *Journal of computational physics*, 143(2):403–425, 1998.

[57] K. Ito and S. S. Ravindran. Reduced basis method for optimal control of unsteady viscous flows. *International Journal of Computational Fluid Dynamics*, 15(2):97–113, 2001.

[58] A. Jerri. *Introduction to Integral Equations with Applications*. A Wiley-Interscience publication. Wiley, 1999.

[59] G.-S. Jiang and C.-W. Shu. Efficient implementation of weighted ENO schemes. *Journal of Computational Physics*, 126(1):202–228, 1996.

[60] I. Kalashnikova, B. van Bloemen Waanders, S. Arunajatesan, and M. Barone. Stabilization of projection-based reduced order models for linear time-invariant

systems via optimization-based eigenvalue reassignment. *Computer Methods in Applied Mechanics and Engineering*, 272:251–270, 2014.

[61] M. Karow, D. Kressner, and F. Tisseur. Structured eigenvalue condition numbers. *SIAM Journal on Matrix Analysis and Applications*, 28(4):1052–1068, 2006.

[62] N.-R. Kevlahan and M. Farge. Vorticity filaments in two-dimensional turbulence: creation, stability and effect. *Journal of Fluid Mechanics*, 346:49–76, 1997.

[63] A. Kolmogoroff. Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Annals of Mathematics. Second Series*, 37(1):107–110, 1936.

[64] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis*, 40(2):492–515, 2002.

[65] S. Lall, P. Krysl, and J. E. Marsden. Structure-preserving model reduction for mechanical systems. *Physica D: Nonlinear Phenomena*, 184(1-4):304–318, 2003.

[66] H. Lamb. On a peculiarity of the wave-system due to the free vibrations of a nucleus in an extended medium. *Proc. Lond. Math. Soc. XXXII*, pages 208–211, 1900.

[67] H. Langtangen and A. Logg. *Solving PDEs in Python: The FEniCS Tutorial I*. Simula SpringerBriefs on Computing. Springer International Publishing, 2017.

[68] I. Markovsky. *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer Publishing Company, Incorporated, 2011.

[69] J. E. Marsden and T. S. Ratiu. *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*, volume 17. Springer Science & Business Media, 2013.

[70] V. Mehrmann and F. Poloni. Doubling algorithms with permuted lagrangian graph bases. *SIAM Journal on Matrix Analysis and Applications*, 33(3):780–805, 2012.

[71] V. Mehrmann and F. Poloni. An inverse-free adi algorithm for computing lagrangian invariant subspaces. *Numerical Linear Algebra with Applications*, 23(1):147–168, 2016.

[72] V. Mehrmann and D. Watkins. Structure-preserving methods for computing eigenpairs of large sparse skew-hamiltonian/hamiltonian pencils. *SIAM Journal on Scientific Computing*, 22(6):1905–1925, 2001.

[73] T. Misumi, M. Nitta, and N. Sakai. Resurgence in sine-Gordon quantum mechanics: exact agreement between multi-instantons and uniform wkb. *Journal of High Energy Physics*, 2015(9):157, Sep 2015.

[74] K. S. Mohamed. *Machine Learning for Model Order Reduction*. Springer, 2018.

[75] Y. Morinishi. Skew-symmetric form of convective terms and fully conservative finite difference schemes for variable density low-mach number flows. *Journal of Computational Physics*, 229(2):276–300, 2010.

[76] Y. Morinishi, T. S. Lund, O. V. Vasilyev, and P. Moin. Fully conservative higher order finite difference schemes for incompressible flow. *Journal of computational physics*, 143(1):90–124, 1998.

[77] Y. Morinishi, S. Tamano, and K. Nakabayashi. A dns algorithm using b-spline collocation method for compressible turbulent channel flow. *Computers & fluids*, 32(5):751–776, 2003.

[78] E. Musharbash, F. Nobile, and T. Zhou. Error analysis of the dynamically orthogonal approximation of time dependent random pdes. *SIAM Journal on Scientific Computing*, 37(2):A776–A810, 2015.

[79] F. Negri, A. Manzoni, and D. Amsallem. Efficient model reduction of parametrized systems by matrix discrete empirical interpolation. *Journal of Computational Physics*, 303:431–454, 2015.

[80] L. Peng and K. Mohseni. Geometric model reduction of forced and dissipative hamiltonian systems. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 7465–7470. IEEE, 2016.

[81] L. Peng and K. Mohseni. Symplectic model reduction of hamiltonian systems. *SIAM Journal on Scientific Computing*, 38(1):A1–A27, 2016.

[82] J. S. Peterson. The reduced basis method for incompressible viscous flow calculations. *SIAM Journal on Scientific and Statistical Computing*, 10(4):777–786, 1989.

[83] A. Pinkus. *N-widths in approximation theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, 1985.

[84] R. V. Polyuga and A. Van der Schaft. Structure preserving model reduction of port-hamiltonian systems by moment matching at infinity. *Automatica*, 46(4):665–672, 2010.

[85] S. Prajna. Pod model reduction with stability guarantee. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 5, pages 5254–5258. IEEE, 2003.

[86] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*. UNITEXT. Springer International Publishing, 2015.

[87] S. S. Ravindran. Adaptive reduced-order controllers for a thermal flow system using proper orthogonal decomposition. *SIAM Journal on Scientific Computing*, 23(6):1924–1942, 2002.

[88] J. Reiss and J. Sesterhenn. A conservative, skew-symmetric finite difference scheme for the compressible navier–stokes equations. *Computers & Fluids*, 101:208–219, 2014.

[89] N. Ripamonti. Energy-preserving model reduction of fluid flows. Master's thesis, EPFL, 2017.

[90] J. W. Robbin and D. A. Salamon. Introduction to differential geometry. 2018.

[91] G. Rozza. Reduced-basis methods for elliptic equations in sub-domains with a posteriori error bounds and adaptivity. *Applied Numerical Mathematics*, 55(4):403–424, 2005.

[92] W. Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1976.

[93] W. Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

[94] T. Ruiner, J. Fehr, B. Haasdonk, and P. Eberhard. A-posteriori error estimation for second order mechanical systems. *Acta Mechanica Sinica*, 28(3):854–862, 2012.

[95] A. Salam and E. Al-Aidarous. Equivalence between modified symplectic gram-schmidt and householder sr algorithms. *BIT Numerical Mathematics*, 54(1):283–302, 2014.

[96] B. Schutz, B. Tapley, and G. H. Born. *Statistical orbit determination*. Academic Press, 2004.

[97] S. Sen, K. Veroy, D. Huynh, S. Deparis, N. C. Nguyen, and A. T. Patera. "natural norm" a posteriori error estimators for reduced basis approximations. *Journal of Computational Physics*, 217(1):37–62, 2006.

[98] B. Sjögreen and H. Yee. On skew-symmetric splitting and entropy conservation schemes for the euler equations. In *Numerical Mathematics and Advanced Applications 2009*, pages 817–827. Springer, 2010.

[99] R. Smith, M. Ellis, G. Xia, V. Sankaran, W. Anderson, and C. Merkle. Computational investigation of acoustics and instabilities in a longitudinal-mode rocket combustor. *AIAA Journal*, 46(11):2659–2673, 2008.

[100] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, MA, fourth edition, 2009.

[101] E. Tadmor. Skew-selfadjoint form for systems of conservation laws. *Journal of Mathematical Analysis and Applications*, 103(2):428–442, 1984.

[102] G. Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Society Providence, 2012.

[103] W. J. Thompson. Fourier series and the gibbs phenomenon. *American journal of physics*, 60(5):425–429, 1992.

[104] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, 1997.

[105] A. van der Schaft. *L_2-gain and passivity techniques in nonlinear control*, volume 218 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag London, Ltd., London, 1996.

[106] A. van der Schaft and D. Jeltsema. Port-hamiltonian systems theory: An introductory overview. *Found. Trends Syst. Control*, 1(2-3):173–378, June 2014.

[107] R. M. Wald. *General relativity*. Chicago Univ. Press, Chicago, IL, 1984.

[108] Q. Wang, J. S. Hesthaven, and D. Ray. Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *Journal of computational physics*, 2018.

[109] D. S. Watkins. On hamiltonian and symplectic lanczos processes. *Linear algebra and its applications*, 385:23–45, 2004.

[110] J. C. Willems. Dissipative dynamical systems. II. Linear systems with quadratic supply rates. *Archive for Rational Mechanics and Analysis*, 45:352–393, 1972.

[111] H. Xu. An svd-like matrix decomposition and its applications. *Linear algebra and its applications*, 368:1–24, 2003.

[112] J. Yu and J. S. Hesthaven. A comparative study of shock capturing models for the discontinuous galerkin method. *No. EPFL-ARTICLE-231188.*, 2017.

[113] Y. Yu, S. Koeglmeier, J. Sisco, and W. Anderson. Combustion instability of gaseous fuels in a continuously variable resonance chamber (CVRC). In *44th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, page 4657, 2008.

BABAK MABOUDI AFKHAM

PERSONAL INFORMATION

Born on 22 March 1989

Nationality Iranian

email babak.maboudi@epfl.ch

phone (M) +41 78 627 46 97

INTERESTS

Research Model Order Reduction.

Other Interests Differential Geometry, Approximation Theory, Uncertainty Quantification, Machine Learning, High-Performance Computing

EDUCATION

Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne-Switzerland

2014-present Ph.D. in Computational Mathematics and Simulation Science
Advisor: Prof. Jan S. Hesthaven
Research topic: Structure-Preserving Model-Reduction

Massachusetts Institute of Technology (MIT), Cambridge-United States of America

2017-2018 Exchange Graduate Student in Aeronautics and Astronautics
Advisor: Prof. Karen Willcox
Research topic: Energy-Preserving Model-Reduction for Euler's Equation

Royal Institute of Technology (KTH), Stockholm-Sweden

2012-2014 M.Sc. in Scientific Computing
Advisor: Prof. Anna-Karin Tornberg
Thesis topic: Simulation of elastic rods with intrinsic curvature and twist immersed in fluid

Sharif University of Technology (SUT), Tehran-Iran

2007-2012 B.Sc. in Theoretical Mathematics
Advisor: Prof. Mohammad Reza Razvan
Thesis topic: Learning Spectral Clustering

AWARDS

2017 The SNSF Doc.Mobility grant, 2017.

2014 The SMC (Stockholm Mathematics Center) award for excellent master thesis, 2014.

2013 KTH tuition fee waiver, 2013.

PUBLICATIONS

2018 Babak Maboudi Afkham, Jan S. Hesthaven, "Structure-Preserving Model-Reduction of Dissipative Hamiltonian System", Journal of Scientific Computing (2018): 1-19

2017 Babak Maboudi Afkham, Jan S. Hesthaven, "Structure-Preserving

Model-Reduction of Parametric Hamiltonian System", SIAM Journal on Scientific Computing 39.6 (2017): A2616-A2644

UNDER REVIEW/PREPARATION WORK

2018 Babak Maboudi Afkham, Nicolò Ripamonti, Qian Wang, Jan Hesthaven, "Conservative Model Order Reduction for Fluid Flow" - Submited to MS&A, Springer.

2018 Babak Maboudi Afkham, Ashish Bhatt, Bernard Haasdonk, Jan S. Hesthaven, "Symplectic Model Reduction with a Weighted Inner Product", under preparation.

TEACHING AND SUPERVISION

2014-2017 Principal Teacher Assistant of Analysis I and II: Holding 8 hours of lecture, Holding Exercise classes,Designing weekly exercise sheets

2017 Co-supervisor of the master thesis: "Energy preserving model reduction of fluid dynamics", Nicolo Ripamonti

2015 Supervisor of the semester project: "Hamiltonian formulation for non-conservative systems", Bozorgmehr Aminian

INVITED TALKS AT INTERNATIONAL CONFERENCES AND WORKSHOPS

2018 MoRePaS 2018 Conference - Nantes, France
Keynote: "Model Order Reduction While Preserving a First Integral"

2016 MORCIP - Workshop on Model Order Reduction for Control & Inverse Problems, EPFL
Invited Speaker: "Structure-Preserving Model Reduction of Hamiltonian Systems"

2016 ALOP - Workshop on Reduced Order Models in Optimization, The University of Trier
Invited Speaker: "Structure-Preserving Model Reduction of Hamiltonian Systems"

LANGUAGES

English (Professional working proficiency), Persian (Mother Tongue), French (Intermediate Proficiency)

HOBBIES

Rock-climbing, Mountaineering (Mount Kilimanjaro 5895m, Mount Damavand 5678m), Distance Running

REFERENCES

Prof. Jan S. Hesthaven
Ecole Polytechnique Fédérale de Lausanne (EPFL)

Prof. Bernard Haasdonk
University of Stuttgart