SYMPLECTIC MODEL REDUCTION OF HAMILTONIAN SYSTEMS

LIQIAN PENG and KAMRAN MOHSENI

Abstract. In this paper, a symplectic model reduction technique, proper symplectic decomposition (PSD) with symplectic Galerkin projection is proposed to achieve computational savings for the simplification of large-scale Hamiltonian systems while preserving the symplectic structure. As an analogy to the classical POD-Galerkin approach, the PSD is designed to build a symplectic subspace to fit empirical data, while the symplectic Galerkin projection constructs a low-order Hamiltonian system on the symplectic subspace. For practical use, we introduce three algorithms for PSD, which are based upon: nonlinear programming, the cotangent lift, and the complex singular value decomposition. The proposed technique has been proved to preserve system energy, volume of flow, and stability. Moreover, PSD can be combined with the discrete empirical interpolation method to reduce the computational cost for nonlinear Hamiltonian systems. Due to these properties, the proposed technique can be better suited for model reduction of hyperbolic PDEs compared to the classical POD-Galerkin approach, especially when long-time integration is needed. The stability, accuracy, and efficiency of the proposed technique are illustrated through numerical simulations of linear and nonlinear wave equations.

Key words. Symplectic model reduction, Hamiltonian system, proper symplectic decomposition (PSD), symplectic Galerkin projection, symplectic structure preservation, stability preservation, symplectic discrete empirical interpolation method (SDEIM)

AMS subject classifications. 65P10, 37M15, 34C20, 93A15, 37J25

1. Introduction. In order to achieve computational savings, model reduction seeks to approximate high dimensional dynamical systems by simpler, lower order ones that can capture the dominant properties. The need for model reduction arises because, in many cases, the direct numerical simulations are often so computationally intensive that they either cannot be performed as often as needed or are more often performed in special circumstances than routinely. See [2] for a survey on the available model reduction methods.

Among these methods, the proper orthogonal decomposition (POD) with Galerkin projection, which was first introduced by Moore [19], has wide applications in electrical power grids [20], structural dynamics [1], and computational-fluid-dynamics-based modeling and control [12, 26], to list a few. As a data-based model reduction method, the POD-Galerkin method involves an offline-online splitting methodology. In the offline stage, direct numerical simulations are carried out to generate empirical data. The POD method is used to compute the optimal approximating subspace to fit the empirical data. A reduced system is then obtained by projecting the original system to this subspace. In the online stage, one can solve the reduced system in the optimal subspace. Recently, many variants of POD-Galerkin were developed to reduce the complexity of evaluating the nonlinear term of the vector field, such as the trajectory piecewise linear and quadratic approximations [23], the missing point estimation [3], the Gappy POD method [27, 29], the empirical interpolation method [4, 10], and the discrete empirical interpolation method (DEIM) [7, 9].

Generally, the classical POD-Galerkin method is not guaranteed to yield a stable
reduced system, even if the original system is stable \cite{21, 23}. The instability of a reduced system is often accompanied by the blowup of system energy and the flow volume. Therefore, when the original large-scale system is conservative, it is preferable to construct a low dimensional reduced system that preserves the geometric structure and allows symplectic integrators. However, much less effort has been expended in the field of geometric model reduction. In the context of a Lagrangian system, Lall et al. \cite{13} show that performing a Galerkin projection on the Euler–Lagrange equation and lifting it to the tangent bundle of the phase space lead to a reduced-order model that preserves Lagrangian structure. In order to reduce the complexity of nonlinear Lagrangian systems, Carlberg et al. combine the Lall’s method with the gappy POD method to derive reduced-order equations \cite{6}, and speedups are obtained for nonlinear Lagrangian systems.

In this paper, we propose a new model reduction method that preserves the symplectic structure underlying the Hamiltonian mechanics, and that ties in with the POD-Galerkin method. Since the Hamiltonian is the Legendre transform of the Lagrangian, both approaches give equivalent equations for the same physical problem. Our motivation to use Hamiltonian approach instead of Lagrangian approach comes from many desirable properties of the symplectic structure of Hamiltonian systems. Our main focus is to develop a basic framework behind symplectic model reduction, which allows us to derive energy preservation, volume preservation, and stability preservation. The proposed technique yields a canonical form of reduced-order Hamiltonian systems which are applicable to the long-time integration. Compared with model reduction algorithms based on Lagrangian mechanics, the proposed technique is easier to apply to some problems. It also provides increased flexibility to construct an optimal subspace that can yield a more accurate solution for the same subspace dimension.

The remainder of this paper is organized as follows. Preliminaries of Hamiltonian systems and symplectic integrators are briefly reviewed in section 2. The symplectic projection, which constructs reduced-order Hamilton’s equations, is presented in section 3 and proved to be energy preserving, volume preserving, and stability preserving. In section 4, three different algorithms are proposed to construct a symplectic matrix, including the nonlinear optimization method, the cotangent lift method, and the complex SVD method. In section 5, the symplectic discrete empirical interpolation method (SDEIM) is developed in order to reduce the complexity of evaluating the nonlinear vector term. Sections 3, 4, 5 respectively associate with the classical Galerkin projection, POD, and DEIM. In section 6, the stability, accuracy, and efficiency of the proposed technique are illustrated through numerical simulations of linear and nonlinear wave equations. Finally, conclusions are offered in section 7.

2. Hamilton’s Equation and Symplectic Integrator. Let an \(n\)-dimensional real vector space \(Q\) be a configuration space. We can select coordinates \((q_1, \ldots, q_n)\) on \(Q\). Let \((p_1, \ldots, p_n)\) be the coordinates of the dual space \(Q^*\). With \(T^*Q = Q \times Q^*\), the pair \((T^*Q, \Omega)\) is a symplectic space, where the cotangent bundle \(T^*Q\) is the phase space and \(\Omega\) is a closed non-degenerate two-form on \(T^*Q\). Moreover, \(\Omega\) takes the following canonical symplectic form, \(\Omega = \sum_{i=1}^{n} dq_i \wedge dp_i\). Throughout this paper, we implicitly assume that \(T^*Q\) is a vector space over the field \(\mathbb{R}\). In this case, \(T^*Q = \mathbb{R}^{2n}\), and for all \(y \in T^*Q\) and \(v_1, v_2 \in T_y(T^*Q)\), we have the following equation

\[
\Omega_y(v_1, v_2) = v_1^T J_{2n} v_2, \quad J_{2n} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix},
\]
where $J_{2n}$ is the Poisson matrix and $I_n$ is the $n \times n$ identity matrix. It is easy to verify that $J_{2n}J_{2n}^T = J_{2n}^T J_{2n} = I_{2n}$, and $J_{2n}J_{2n} = J_{2n}^T J_{2n}^T = -I_{2n}$.

Let $H : T^*Q \to \mathbb{R}$ denote a Hamiltonian function. The time evolution of the Hamiltonian system is defined by:

$$\dot{q} = \nabla_p H(q, p), \quad \dot{p} = -\nabla_q H(q, p),$$

where $q \in Q$, and $p \in Q^*$. We abstract this formulation by introducing the phase space variable $y = [q; p]$\footnote{The notations $[q, p]$ and $[q; p]$ are the same as the corresponding functions in MATLAB.} and the abstract Hamiltonian differential equation

$$\dot{y} = X_H(y),$$

where $X_H(y) := J_{2n} \nabla_y H(y)$ is the Hamiltonian vector field. The flow $\Psi_t$ of $X_H$ is a symplectomorphism, meaning that it conserves the symplectic two-form ($\Omega \circ \Psi_t = \Omega$), the system Hamiltonian ($H \circ \Psi_t = H$) and the volume of flow ($\Theta \circ \Psi_t = \Theta$), where $\Theta := dq_1 \wedge \ldots \wedge dq_n \wedge dp_1 \wedge \ldots \wedge dp_n$. We recommend readers to refer \cite{15} for more details of these fundamental facts of symplectic geometry.

Symplectic integrators are numerical schemes for solving the Hamiltonian system, while preserving the underlying symplectic structure. If the symplectic structure is preserved, then the flow volume is automatically conserved during the symplectic time integration. Moreover, symplectic integrators conserve the Hamiltonian with a slightly perturbed “energy drift”. By virtue of these advantages, symplectic integrators have been widely applied to the long-time integrations of the molecular dynamics, discrete element methods, the accelerator physics, and the celestial mechanics \cite{11}.

The simplest symplectic scheme for the numerical integration of a finite-dimensional ODE system is given by

$$q^{j+1} = q^j + \delta t \nabla_p H(q^{j+1}, p^j); \quad p^{j+1} = p^j - \delta t \nabla_q H(q^{j+1}, p^j),$$

or

$$q^{j+1} = q^j + \delta t \nabla_p H(q^j, p^{j+1}); \quad p^{j+1} = p^j - \delta t \nabla_q H(q^j, p^{j+1}),$$

where $\delta t$ is the unit step for time integration. The above methods are implicit for general Hamiltonian systems. For separable $H(q, p) = T(p) + U(q)$, however, both variants turn out to be explicit. If the implicit midpoint rule is applied, then a second order symplectic scheme is obtained:

$$y^{j+1} = y^j + \delta t J_{2n} \nabla_y H \left( \frac{y^{j+1} + y^j}{2} \right).$$

Most of the usual numerical methods, like the primitive Euler scheme and the classical Runge–Kutta scheme, are not symplectic integrators. A comprehensive review of symplectic integrators for Hamiltonian ODEs could be found in \cite{11, 16}, the extension for Hamiltonian PDEs is reviewed by Bridges and Reich \cite{5}.

3. Symplectic Projection. The proposed symplectic model reduction technique in the paper takes advantage of empirical data to construct a low-dimensional system, while simultaneously preserving the underlying symplectic structure. Specifically, if the original system is Hamiltonian, the reduced system remains Hamiltonian, but with significantly fewer dimensions.

\footnote{The volume of flow is also known as Liouville measure.}
3.1. Definition of symplectic projection. Let \((T^*\mathcal{R}, \omega)\) be a symplectic vector space with \(\dim(\mathcal{R}) = k\) for some \(k \leq n\). By choosing coordinates \((r_1, \ldots, r_k)\) for \(\mathcal{R}\) and \((s_1, \ldots, s_k)\) for \(\mathcal{R}^*\), the canonical symplectic form \(\omega\) on \(T^*\mathcal{R}\) has the following form \(\omega = \sum_{i=1}^{k} dr_i \wedge ds_i\).

**Definition 3.1.** Let \((T^*\mathcal{Q}, \Omega)\) and \((T^*\mathcal{R}, \omega)\) be two symplectic vector spaces. A **symplectic lift** is a linear mapping \(\phi : T^*\mathcal{R} \to T^*\mathcal{Q}\) that satisfies
\[
\omega_\phi(u_1, u_2) = \Omega_{\phi(x)}(\phi_\ast u_1, \phi_\ast u_2),
\]
for all \(x \in T^*\mathcal{R}\) and \(u_1, u_2 \in T_x(T^*\mathcal{R})\), where \(\phi_\ast : T_x(T^*\mathcal{R}) \to T_{\phi(x)}(T^*\mathcal{Q})\) denotes the pushforward by \(\phi\).

Using the above coordinates for \(T^*\mathcal{R}\) and \(T^*\mathcal{Q}\), the symplectic lift \(\phi : x \mapsto y\) can be written in matrix form
\[
y = \phi(x) = Ax.
\]
By identifying \(T^*\mathcal{R}\) and \(T^*\mathcal{Q}\) with \(\mathcal{R}\) and \(\mathcal{Q}\) with their corresponding tangent spaces, it immediately follows that \(\phi_\ast(u_1) = Au_1\) and \(\phi_\ast(u_2) = Au_2\). Since \(u_1\) and \(u_2\) are arbitrary vectors in \(T_x(T^*\mathcal{R})\), (3.2) is equivalent to
\[
A^T J_{2n} A = J_{2k},
\]
where superscript \(T\) represents the matrix transpose. A \(2n \times 2k\) matrix \(A\) is called **symplectic** if \(A \in Sp(2k, \mathbb{R}^{2n})\), where
\[
Sp(2k, \mathbb{R}^{2n}) := \{ A \in \mathbb{R}^{2n \times 2k} | A^T J_{2n} A = J_{2k} \}.
\]
Since \(J_{2n}\) and \(J_{2k}\) are nonsingular, \(3.3\) requires that \(k \leq n\) and \(\text{rank}(A) = 2k\). Especially, when \(k = n\), \(A\) is a square matrix and \(Sp(2n, \mathbb{R}^{2n})\) forms a symplectic group, denoted by \(Sp(\mathbb{R}^{2n})\). However, for model reduction purposes, we are more interested for the case \(k \ll n\); we will implicitly assume \(k \leq n\) throughout the rest of this paper.

In mathematics, \(Sp(2k, \mathbb{R}^{2n})\) denotes the **symplectic Stiefel manifold**. It may also be defined as a homogeneous space for the action of a symplectic group. The symplectic group \(Sp(\mathbb{R}^{2n})\) acts transitively on \(Sp(2k, \mathbb{R}^{2n})\), while the stabilizer subgroup of \(A\) is a subgroup of \(Sp(\mathbb{R}^{2n})\) isomorphic to \(Sp(\mathbb{R}^{2n-2k})\). Therefore, \(Sp(2k, \mathbb{R}^{2n}) \cong Sp(\mathbb{R}^{2n})/Sp(\mathbb{R}^{2n-2k})\). Since \(\dim(\text{dim}(Sp(\mathbb{R}^{2n})) = n(2n+1)\), the dimension of \(Sp(2k, \mathbb{R}^{2n})\) is obtained by \(\dim(\text{dim}(Sp(\mathbb{R}^{2n})) - \dim(\text{dim}(Sp(\mathbb{R}^{2n-2k})))\), which equals \(4nk - k(2k - 1)\).

**Definition 3.2.** The **symplectic inverse** of a real matrix \(A \in \mathbb{R}^{2n \times 2k}\), denoted as \(A^+\), is defined by
\[
A^+ = J_{2k}^T A^T J_{2n}.
\]
Although \(A^+\) is not equal to the Moore-Penrose pseudoinverse of \(A\), \(A^+\) has several interesting properties, as stated in the following two lemmas. Using the definition of \(A^+\), it is straightforward to verify Lemma 3.3.

**Lemma 3.3.** Suppose \(A \in \mathbb{R}^{2n \times 2k}\) and \(A^+\) is the symplectic inverse of \(A\). Then,
\[
A = (A^+)^+, \quad (A^+)^T, \quad A = (((A^+)^T)^T)^T.
\]
A\textsuperscript{T} J_{2n} = J_{2k} A\textsuperscript{T}.

\textbf{Lemma 3.4.} Suppose $A \in \mathbb{R}^{2n \times 2k}$ and $A^\dagger$ is the symplectic inverse of $A$. Then the following are equivalent:
(a) $A \in Sp(2k, \mathbb{R}^{2n})$.
(b) $(A^\dagger)^T \in Sp(2k, \mathbb{R}^{2n})$.
(c) $A^\dagger A = I_{2k}$.

\textit{Proof.} (a) $\Rightarrow$ (b) : Replacing $A$ by $A^\dagger$ yields
\begin{equation}
A^\dagger J_{2n}(A^\dagger)^T = (J_{2k}^T A^T J_{2n})J_{2n}(J_{2n}^T A J_{2k}) = J_{2k}^T A^T (J_{2n} J_{2n} J_{2k}^T) A J_{2k} = J_{2k}^T A^T J_{2n} A J_{2k}.
\end{equation}

In addition, $A \in Sp(2k, \mathbb{R}^{2n})$ implies that $A^T J_{2n} A = J_{2k}$. It follows that
\begin{equation}
A^\dagger J_{2n}(A^\dagger)^T = J_{2k}^T (A^T J_{2n} A) J_{2k} = J_{2k}^T J_{2k} J_{2k} = J_{2k}.
\end{equation}

Since $((A^\dagger)^T)^T = A^\dagger$, we have $(A^\dagger)^T \in Sp(2k, \mathbb{R}^{2n})$.

(b) $\Rightarrow$ (c) : Substituting $A$ by $A^\dagger$,
\begin{equation}
A^\dagger A = A^\dagger ((A^\dagger)^T)^T = A^\dagger (J_{2k}^T A^T J_{2n})^T
= A^\dagger J_{2n} (A^\dagger)^T J_{2k} = -(A^\dagger J_{2n} (A^\dagger)^T) J_{2k}
\end{equation}

Since $(A^\dagger)^T \in Sp(2k, \mathbb{R}^{2n})$, (3.8) holds. It follows that
\begin{equation}
A^\dagger A = -J_{2k} J_{2k} = I_{2k}.
\end{equation}

(c) $\Rightarrow$ (a) : Plugging (3.4) into (3.9), we have $J_{2k}^T A^T J_{2n} A = I_{2k}$. Left multiplying $J_{2k}$ on both sides of this equation yields $A^T J_{2n} A = J_{2k}$.

\textbf{Definition 3.5.} Let $x \in T^\ast \mathcal{R}$, and $y \in T^\ast \mathcal{Q}$. Using the standard coordinates, a linear mapping $\pi : y \mapsto x$ is a \textit{symplectic projection} if there exist a symplectic matrix $A \in Sp(2k, \mathbb{R}^{2n})$, such that
\begin{equation}
x = A^\dagger y.
\end{equation}

By (3.9) in Lemma 3.4, $\pi \circ \phi$ is the identity map on $T^\ast \mathcal{R}$. Now suppose $y = Ax$. Taking the time derivative of (3.10), using (2.2), and the chain rule, the time evolution of $x$ is given by
\begin{equation}
\dot{x} = A^\dagger \dot{y} = A^\dagger J_{2n} \nabla_y H(y) = A^\dagger J_{2n} (A^\dagger)^T \nabla_x H(Ax).
\end{equation}

A necessary and sufficient condition for the last equation to be Hamiltonian is that
\begin{equation}
A^\dagger J_{2n} (A^\dagger)^T = J_{2k}.
\end{equation}

By Lemma 3.4, this is equivalent to $A^T J_{2n} A = J_{2n}$, i.e., $A \in Sp(2k, \mathbb{R}^{2n})$.

\textbf{Definition 3.6.} The \textit{symplectic Galerkin projection}, or \textit{symplectic projection}, of a $2n$-dimensional Hamilton’s equation $\dot{y} = J_{2n} \nabla_y H(y)$, with an initial condition $y(0) = y_0$ is given by a $2k$-dimensional ($k \leq n$) system
\begin{equation}
\dot{x} = J_{2k} \nabla_x \tilde{H}(x); \quad x_0 = A^\dagger y_0,
\end{equation}
where $\tilde{H} := H \circ A$ is the reduced Hamiltonian function, $A \in \mathbb{R}^{2n \times 2k}$ is a symplectic matrix, and $A^+ \in \mathbb{R}^{2k \times 2n}$ is the symplectic inverse of $A$.

A Hamiltonian system is linear if $H(y) = \frac{1}{2} y^T L y$, where $L$ is a $2n \times 2n$ real symmetric matrix. Let $K := J_{2n} L$, the linear Hamilton’s equation can be written as

$$\dot{y} = J_{2n} L y = \tilde{K} y. \tag{3.12}$$

A matrix of the form $K = J_{2n} L$, where $L$ is symmetric, is called a Hamiltonian matrix. Moreover, the set of all $2n \times 2n$ Hamiltonian matrices, denoted by $\mathfrak{sp}(\mathbb{R}^{2n})$, is a Lie algebra [17]. The fundamental matrix solution to (3.12) is given by

$$y(t) = e^{Kt} y_0.$$ \tag{3.13}

Since $\exp(Kt)$ satisfies $(\exp(Kt))^T J_{2n} \exp(Kt) = J_{2n}$, we have $\exp(tK) \in \mathfrak{sp}(\mathbb{R}^{2n})$, which means the matrix exponential of a Hamiltonian matrix is symplectic. Conversely, the logarithm of a square symplectic matrix is Hamiltonian, see reference [17] for the proof.

Applying the symplectic projection on the linear system (3.12) gives

$$\dot{x} = A^+ \dot{y} = A^+ K y = A^+ K A x = \tilde{K} x. \tag{3.14}$$

where $\tilde{K} := A^+ K A$ is the reduced linear operator. Using (3.7) in Lemma 3.3, we obtain

$$\tilde{K} = A^+ (J_{2n} L) A = J_{2k} (A^T L A) = J_{2k} \tilde{L}. \tag{3.15}$$

Since $\tilde{L} = A^T L A$ is symmetric, $\tilde{K}$ is also Hamiltonian. In fact, (3.14) can also be obtained by plugging $H(x) = \frac{1}{2} (A x)^T L (A x)$ into (3.11).

Generally, a reduced system constructed by the standard Galerkin projection does not have to be Hamiltonian, consequently energy and volume are not conserved during the time evolution. In contrast, the symplectic projection presented here preserves symplectic structure of Hamiltonian systems, therefore energy and volume are also preserved.

### 3.2. Preservation of Energy, Volume, and Stability

Let $\Psi_t$ and $\psi_t$ respectively denote the Hamiltonian flow of the original system (2.2) and the reduced system (3.11). By definition, we have $y(t) = \Psi_t(y_0)$ and $x(t) = \psi_t(x_0)$. The approximating solution in the original coordinate system is given by $\tilde{y}(t) = A x(t)$. Alteratively, we can write it as $\tilde{y}(t) = \tilde{\Psi}_t(y_0)$, where $\tilde{\Psi}_t = A \circ \psi_t \circ A^+$. Since the energy of the full Hamiltonian system is preserved, we have

$$H(y(t)) = H \circ \Psi_t(y_0) = H(y_0). \tag{3.16}$$

Analogously, energy conservation of the reduced Hamiltonian system gives

$$H(\tilde{y}(t)) = (H \circ A)x(t) = \tilde{H}(\psi_t x_0) = (\tilde{H} \circ \psi_t)(x_0) = \tilde{H}(x_0) = H(A x_0) = H(A A^+ y_0). \tag{3.17}$$

Let $\Delta H(t) := |H(\tilde{y}(t)) - H(y(t))|$ denote the energy discrepancy between the state $y(t)$ and its approximation, $\tilde{y}(t)$, based on a reduced system. If $\Delta H(t) = 0$ holds for any $t$ in the interested time domain, we say the reduced system is energy preserving. Equations (3.16) and (3.17) indicate that, for the symplectic projection, $\Delta H(t)$ is determined by the initial condition $y_0$ and the basis matrix $A$, i.e.,

$$\Delta H(t) = H(y_0) - H(A A^+ y_0). \tag{3.18}$$
If \( y_0 \in \text{Range}(A) \), then there exists a \( x_0 \in \mathbb{R}^{2k} \) such that \( y_0 = Ax_0 \). It follows that 
\[
AA^+y_0 = A(A^+A)x_0 = Ax_0 = y_0.
\]
It follows that the projection error, 
\[
r_0 := y_0 - AA^+y_0,
\]
of \( y_0 \) vanishes. Meanwhile, \( \Delta H(t) = 0 \) is satisfied.

If \( y_0 \notin \text{Range}(A) \), the energy error \( \Delta H(t) \) could be directly computed by \( 3.18 \) even before the reduced system is evolved. If \( H \) is continuously differentiable, we have 
\[
\Delta H(t) \approx DH(y_0) r_0 \text{ as the first order approximation of } 3.18.
\]
As long as \( \|r_0\| \ll 1 \), \( \Delta H(t) \ll 1 \) is obtained. Otherwise, we can always extend \( A \) to a larger symplectic matrix \( A_{ext} \) such that the symplectic projection can generate a reduced system with the energy-preservation property. Specifically, suppose \( A = [A_1, A_2] \) for \( A_1, A_2 \in \mathbb{R}^{2n \times k} \). If \( y_0 \notin \text{Range}(A) \), we must have \( r_0 \neq 0 \). Otherwise, by \( 3.19 \), we obtain 
\[
y_0 = A(A^+y_0),
\]
which breaks the assumption. Thus, the unit vector \( \hat{r}_0 = r_0/\|r_0\| \) is well-defined. One possible extension of \( A \) can be constructed by
\[
A_{ext} = [A_1, \hat{r}_0, A_2, J_{2n}^T \hat{r}_0].
\]
It is straightforward to verify that \( A_{ext}^T A_{ext} J_{2n} A_{ext} = J_{2k+2} \), and \( y_0 - A_{ext} A_{ext}^+ y_0 = 0 \). The last equation means \( y_0 \in \text{Range}(A_{ext}) \). Thus, we have 
\[
H(y_0) - H(A_{ext} A_{ext}^+ y_0) = 0.
\]

As in section 2, let \( \Theta \) denote the volume of flow in \( \mathbb{R}^{2n} \). Analogously, if \( |\Theta(y(t)) - \Theta(\hat{y}(t))| = 0 \) holds for any \( t \) in the interested time domain, we say a reduced system is volume preserving. Because of volume conservation, the reduced Hamiltonian system \( 3.11 \) preserves the volume of the original system if \( y_0 \in \text{Range}(A) \). Otherwise, we can construct a larger symplectic matrix \( A_{ext} \) by \( 3.20 \) such that \( y_0 \in \text{Range}(A_{ext}) \). Then, the new system based on \( A_{ext} \) preserves the volume.

Moreover, the energy conservation also leads to the stability preservation for many Hamiltonian systems. Let \( S \) denote an open set of \( \mathbb{R}^{2n} \) that contains \( y_0 \), and let \( \partial S \) denote the boundary of \( S \). Moreover, we assume \( y_0 \in \text{Range}(A) \), and the initial condition of the reduced system is given by \( x_0 = A^+y_0 \).

**Theorem 3.7.** Consider the Hamiltonian system in \( 3.22 \) with the initial condition \( y_0 \in \mathbb{R}^{2n} \). If there exists a bounded neighbourhood \( S \) of \( y_0 \) in \( \mathbb{R}^{2n} \) such that 
\[
H(y_0) < H(y), \text{ or } H(y_0) > H(y), \text{ for all } y \in \partial S,
\]
then both the original system and the reduced system constructed by the symplectic projection are bounded for all \( t \in \mathbb{R} \).

**Proof.** We first prove that the statement is true for the case that \( H(y_0) < H(y) \) for all \( y \in \partial S \). Let \( E = \min\{H(y) : y \in \partial S\} \), so \( H(y_0) < E \). By \( 3.10 \), we have 
\[
H(\Psi_t(y_0)) = H(y_0) < E.
\]
It follows that \( \Psi_t(y_0) \in S \) for all \( t \), because if not, there is a time \( t_1 \) when \( \Psi_t(y_0) \in \partial S \), and \( H(\Psi_t(y_0)) \geq E \), a contradiction.

Let \( S_A = S \cap \text{Range}(A) \). Since \( S \) is a bounded open set in \( \mathbb{R}^{2n} \), \( S_A \) is also open in \( \text{Range}(A) \) and bounded. Moreover, \( \partial S_A = \partial S \cap \text{Range}(A) \). Thus, \( H(y_0) < H(y) \) for all \( y \in S_A \). By the same argument in the last paragraph, we must have \( \Psi_t(y_0) \in S_A \) for all \( t \), which means that the reduced system constructed by the symplectic projection is also bounded.

Finally, if \( H(y_0) < H(y) \) is exchanged with \( H(y_0) > H(y) \) for all \( y \in S \), the conclusion still holds, because \( -H(y_0) < -H(y) \) and \( -H(\Psi_t(y_0)) \) is a conserved quantity. \( \square \)

The symplectic projection is analogous to the standard Galerkin projection, both of which construct reduced equations in some low-order subspaces. However, the symplectic projection yields a symplectic system by \( 3.11 \) while the standard Galerkin projection generally destroys the symplectic structure. Evolving the system \( 3.11 \) by
a symplectic integrator preserves the system energy, the volume of flow, and the stability of the system. Furthermore, if all the empirical data points approximately lie on the subspace spanned by $A$, (3.11) provides an accurate approximation for the original system. In the next subsection, we shall discuss the proper symplectic decomposition (PSD) to construct the symplectic matrix $A$. This approach is an analogy to the POD method that constructs the orthonormal basis.

4. Proper Symplectic Decomposition (PSD). Let $q(t_i), p(t_i) \in \mathbb{R}^n$ ($i = 1, \ldots, N$) denote empirical data described by canonical coordinates. Each data point $y(t_i) = [q(t_i); p(t_i)]$ is a point in the phase space. Define a snapshot matrix in $\mathbb{R}^{2n \times N}$, 

$$Y := [y(t_1), \ldots, y(t_N)].$$

By (3.10), the symplectic projection of $Y$ onto a low-order subspace is given by $X = A^+ Y$, where $X = [x(t_1), \ldots, x(t_N)] \in \mathbb{R}^{2k \times N}$, and $x(t_i) = [r(t_i); s(t_i)]$ is the corresponding reduced state of $y(t_i)$. The same projection of $Y$ in the original coordinates is given by 

$$\tilde{Y} = AX = AA^+ Y.$$ 

In this section, we propose three different algorithms to construct the symplectic matrix $A$: these are nonlinear programming, the cotangent lift, and the complex SVD.

4.1. Nonlinear Programming. Suppose a symplectic matrix $A$ minimizes the projection error for representing $Y$ onto the column subspace in a least squares sense. The Frobenius norm $\| \cdot \|_F$ is used to measure the difference between $Y$ with its projection $\tilde{Y}$. This problem can be expressed with the following programming problem in the standard form:

$$\begin{align*}
\text{minimize} & \quad \|Y - AA^+ Y\|_F \\
\text{subject to} & \quad A^T J_{2n} A = J_{2k}.
\end{align*}$$

The column space of $A$ is optimal for model reduction since no other symplectic subspace can represent better snapshot approximation in the same dimension. Plugging (3.3) into the objective function yields

$$\|Y - AA^+ Y\|_F = \text{tr}[(Y^T - (AA^+ Y)^T)(Y - AA^+ Y)]$$

$$= \text{tr}[Y^T Y - Y^T AJ_{2k}^T A^T J_{2n} Y - Y^T J_{2n}^T AJ_{2k} A^T Y + Y^T J_{2n}^T AJ_{2k} A J_{2k} A^T J_{2n} Y].$$

Notice that the above expression has a fourth-order term in $A$, the optimization problem can only be solved iteratively by a nonlinear programming method.

4.2. Cotangent Lift. In this section, we propose a SVD-based algorithm to directly construct a symplectic matrix. The idea is to search the optimal matrix, $A_1$, in a subset of $Sp(2k, \mathbb{R}^{2n})$, such that all the empirical data approximately lies on the range of $A_1$. Especially, we define a set $\mathcal{M}_1(2n, 2k)$ by

$$\mathcal{M}_1(2n, 2k) := Sp(2k, \mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} : \Phi \in \mathbb{R}^{n \times k} \right\}.$$ 

If $A_1 \in \mathcal{M}_1(2n, 2k)$, $A_1 = \text{diag}(\Phi, \Phi)$ for some $\Phi \in \mathbb{R}^{n \times k}$. Plugging it into (3.3), we obtain $A_1 \in Sp(2k, \mathbb{R}^{2n})$ iff $\Phi^T \Phi = I_k$, which implies that $\Phi$ is an element of the
Stiefel manifold $V_k(\mathbb{R}^n)$. It follows that,

$$
M_1(2n, 2k) = \left\{ \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \middle| \Phi \in V_k(\mathbb{R}^n) \right\}.
$$

Then, $M_1(2n, 2k) \cong V_k(\mathbb{R}^n)$. As a result, $M_1(2n, 2k)$ is a submanifold of $Sp(2k, \mathbb{R}^{2n})$ with dimension $nk - k(k + 1)/2$.

Let $\mathcal{R}$ and $Q$ denote two vector spaces; $\dim(\mathcal{R}) = k$, $\dim(Q) = n$, and $k \leq n$. Suppose $f : \mathcal{R} \to Q$ and $g : Q \to \mathcal{R}$ are linear mappings and satisfy $g \circ f = id_{\mathcal{R}}$. Let $f_* : T\mathcal{R} \to T^*Q$ denote the pushforward, or the tangent lift by $f$. Let $\langle \cdot, \cdot \rangle$ denote the inner product. The pullback, or the cotangent lift, $f^* : T^*Q \to T^*\mathcal{R}$, of $f$ is a linear mapping that satisfies

$$
\langle f^*(y), w \rangle_{\mathcal{R}} = \langle y, f_*(w) \rangle_{Q}
$$

for any $y \in Q$, $r = g(y) \in \mathcal{R}$, $w \in T_r\mathcal{R}$, and $y \in T^*_rQ$. In the standard coordinates, we have $f(r) = \Phi r$ for some $\Phi \in \mathbb{R}^{n \times k}$. By identifying $\mathcal{R}$ with its tangent space, the tangent lift of $f$ gives $f_*(w) = A_1w$, where $A_1 = \text{diag}(\Phi, \Phi)$. By (4.6), the cotangent lift $y \mapsto x$ can be written as

$$
x = f^*(y) = A_1^T y.
$$

By plugging $A_1 = \text{diag}(\Phi, \Phi)$ into (4.6), it is easy to verify that $A_1^T = \text{diag}(\Phi^T, \Phi^T) = A_1^T$, which means the cotangent lift essentially yields a special symplectic projection $A_1^T$. It follows that $A_1^T A_1 = A_1^T \text{T} A_1 = I_{2k}$, i.e., $A_1 \in V_{2k}(\mathbb{R}^{2n})$. On the other hand, for any $A_1 = \text{diag}(\Phi, \Phi) \in V_{2k}(\mathbb{R}^{2n})$, $\Phi^T \Phi = I_k$ is satisfied. Thus, $A_1 \in Sp(2k, \mathbb{R}^{2n})$. Therefore, an equivalent definition of $M_1(2n, 2k)$ is given by

$$
M_1(2n, 2k) = V_{2k}(\mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \middle| \Phi \in \mathbb{R}^{n \times k} \right\}.
$$

It should be mentioned that in [13], a tangent lift method is used to construct a reduced Euler-Lagrange equation to preserve the Lagrangian structure of the original system. Specifically, a POD basis matrix $\Phi$ can be constructed by the SVD of a snapshot matrix $[q(t_1), \ldots, q(t_N)]$ in the configuration space. The original Lagrangian $L(q, \dot{q})$ in the tangent bundle $TQ$ is approximated by $\tilde{L}(r, \dot{r}) = L(\Phi r, \Phi \dot{r})$ in $T\mathcal{R}$. Then, a reduced system is obtained by solving the Euler-Lagrange equation of $\tilde{L}(r, \dot{r})$. Since the Lagrangian mechanics is equivalent to the Hamiltonian mechanics, the method in [13] is essentially equivalent to the cotangent lift algorithm in this paper. However, $q(t)$ and $\dot{q}(t)$ have the same status in the phase space, and a Hermite subspace [22] that directly constructed on the $TQ$ (rather than $Q$ itself) leads to a better approximation of the original Hamiltonian system. An extended data ensemble that contains both $q$ and $\dot{q}$ has already been used for the online manifold learning based on the subspace iteration using reduced systems [21].

The same idea applies for the approximations in the cotangent bundle $T^*Q$. In order to obtain an accurate reduced system, the range of $\Phi$ should approximately fit for both $q$ and $p$. As Algorithm 1 indicates, $\Phi$ can be computed by the SVD of an extended snapshot matrix $X \in \mathbb{R}^{n \times 2N}$, defined by

$$
X := [q(t_1), \ldots, q(t_N), p(t_1), \ldots, p(t_N)].
$$

**Theorem 4.1.** Suppose $Y \in \mathbb{R}^{2n \times N}$ is the snapshot matrix defined by (4.7). The symplectic matrix $A_1$ constructed by Algorithm 1 is the optimal solution that
minimizes the least square error for representing $Y$ onto the column subspace within the manifold $\mathcal{M}_1(2n, 2k)$.

**Proof.** Similar to (4.3), we can express the optimization problem as:

$$
\begin{align*}
\text{minimize} & \quad \|Y - A_1 A_1^T Y\|_F \\
\text{subject to} & \quad A_1 \in \mathcal{M}_1(2n, 2k).
\end{align*}
$$

Let $Y_1 := [q(t_1), \ldots, q(t_N)]$, and $Y_2 := [p(t_1), \ldots, p(t_N)]$. By definition, $Y = [Y_1; Y_2]$, and $X = [Y_1, Y_2]$. Since $A_1 \in \mathcal{M}_1(2n, 2k)$, we have $A_1 = \text{diag}(\Phi, \Phi)$ with $\Phi^T \Phi = I_k$.

Plugging them into the objective function yields

$$
\|Y - A_1 A_1^T Y\|_F = \left\| \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \begin{bmatrix} \Phi^T & 0 \\ 0 & \Phi^T \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \right\|_F = \left\| \begin{bmatrix} (I_n - \Phi \Phi^T) Y_1 \\ (I_n - \Phi \Phi^T) Y_2 \end{bmatrix} \right\|_F.
$$

Thus, $\Phi$ can be directly solved by the truncated SVD of $X$,

$$
X \approx \Phi \Lambda M^T,
$$

where the matrix $\Lambda$ is a $k \times k$ diagonal matrix with nonnegative real numbers on the diagonal, and $\Phi$ and $M$ are real matrices and satisfy $\Phi^T \Phi = M^T M = I_k$. Thus, the symplectic matrix $A_1$ constructed by Algorithm 1 is the optimal solution for the optimization problem (4.9). $\blacksquare$

Similar to the standard POD(SVD) method, the projection error of $Y$ can be determined by the truncated singular values of $X$. As a result, there exists a global error bound for the reduced system constructed by the symplectic projection.

**4.3. Complex SVD.** The cotangent lift method may still keep some redundant dimensions. For instance, consider the case that $p$ has very rapid oscillation in every possible directions while $q$ is moving near a straight line. In this case, the trajectory of $p$ needs a relatively large dimension to approximate while the trajectory of $q$ could be approximately embedded in a one-dimensional space. However, since $A_1^T$ projects the original system onto a $2k$-dimensional space, $k - 1$ modes in $\mathcal{R}$ are redundant in describing $q$. As a result, a block-diagonal matrix might not be the best option for a basis matrix.

If we use $q(t) + ip(t)$ to describe the solution trajectory in the phase space, we can construct a complex snapshot matrix $Z \in \mathbb{C}^{n \times N}$ by

$$
Z := [q(t_1) + ip(t_1), \ldots, q(t_N), +ip(t_N)].
$$

By definition, we have $Z = Y_1 + iY_2$. The orthonormal basis matrix $U \in \mathbb{C}^{n \times k}$ is constructed to minimize the projection error of $Z$ onto the column subspace of $U$. 

---

**Algorithm 1 Cotangent Lift**

**Require:** An empirical data ensemble $\{q(t_i), p(t_i)\}_{i=1}^N$.

**Ensure:** A symplectic matrix $A_1$ in the block-diagonal form.

1. Construct an extended snapshot matrix $X$ as (4.5).
2. Compute the SVD of $X$ to obtain the basis matrix $\Phi$.
3. Construct the symplectic matrix $A_1 = \text{diag}(\Phi, \Phi)$.
Specifically, it is the solution of the following optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \| Z - U U^H Z \|_F \\
\text{subject to} & \quad U^H U = I_k.
\end{align*}
\]

Here \( U^H \) is the conjugate transpose of \( U \). In fact, \( U \) can be obtained by the truncated SVD of \( Z \),

\[
Z \approx U \Sigma V^H,
\]

where the matrix \( \Sigma \) is a \( k \times k \) diagonal matrix with nonnegative real numbers on the diagonal, and \( U \) and \( V \) are complex matrices and satisfy \( U^H U = V^H V = I_k \). Let \( V_k(\mathbb{C}^n) \) denote the Stiefel manifold in \( \mathbb{C}^n \) and write its element \( U \in V_k(\mathbb{C}^n) \) in the form \( U = \Phi + i\Psi \), where \( \Phi, \Psi \in \mathbb{R}^{n \times k} \). We define a mapping \( \mathcal{A} : V_k(\mathbb{C}^n) \rightarrow \mathbb{R}^{2n \times 2k} \) by the formula

\[
\mathcal{A}(\Phi + i\Psi) = \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix},
\]

where \( h(U) \in \mathbb{R}^{2n \times 2k} \) is a block matrix.

**Lemma 4.2.** The mapping \( \mathcal{A} \) is injective. The image of \( \mathcal{A} \) is equal to \( M_2(2n, 2k) \), where

\[
M_2(2n, 2k) := Sp(2k, \mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix} \mid \Phi, \Psi \in \mathbb{R}^{n \times k} \right\}.
\]

**Proof.** It follows from \( \mathcal{A} \)'s definition that it is injective. If \( \Phi + i\Psi \in V_k(\mathbb{C}^n) \), then

\[
(\Phi + i\Psi)^H(\Phi + i\Psi) = I_k,
\]

which is equivalent to

\[
\Phi^T \Phi + \Psi^T \Psi = I_k, \quad \Phi^T \Psi = \Psi^T \Phi.
\]

Let \( A_2 = \mathcal{A}(\Phi + i\Psi) \). Using (4.13), it is easy to verify that \( A_2^T J_{2n} A_2 = J_{2k} \). Thus, \( A_2 \in M_2(2n, 2k) \), i.e., \( \mathcal{A}(V_k(\mathbb{C}^n)) \subseteq M_2(2n, 2k) \).

Conversely, if \( A_2 \in M_2(2n, 2k) \), then \( A_2^T J_{2n} A_2 = J_{2k} \). Moreover, we can write \( A_2 = [\Phi, -\Psi, \Psi, \Phi] \) for some \( \Phi, \Psi \in \mathbb{R}^{n \times k} \). Plugging it into \( A_2^T J_{2n} A_2 = J_{2k} \) gives (4.16). It follows that, \((\Phi + i\Psi)^H(\Phi + i\Psi) = I_k \). As a result, \( \Phi + i\Psi \in V_k(\mathbb{C}^n) \), and \( \mathcal{A}^{-1}(M_2(2n, 2k)) \subseteq V_k(\mathbb{C}^n) \).

**Algorithm 2** Complex SVD

**Require:** An empirical data ensemble \( \{q(t_i), p(t_i)\}_{i=1}^N \).

**Ensure:** A symplectic matrix \( A_2 \) in the block form.

1. Construct a complex snapshot matrix \( Z \) as (4.11).
2. Compute the SVD of \( Z \) to obtain the basis matrix \( \Phi + i\Psi \).
3. Construct the symplectic matrix \( A_2 = [\Phi, -\Psi; \Psi, \Phi] \).

Lemma 4.2 implies that a symplectic matrix \( A_2 \) can be constructed through the mapping \( \mathcal{A} \). Algorithm 2 outlines the procedure. Since both \( \mathcal{A} \) and \( \mathcal{A}^{-1} \) are smooth mappings, \( M_2(2n, 2k) \cong V_k(\mathbb{C}^n) \), and \( M_2(2n, 2k) \) is a submanifold of \( Sp(2k, \mathbb{R}^{2n}) \) with dimension \( 2nk - k^2 \).
Moreover, by substituting $A_2 = [\Phi, \Psi, -\Psi, \Phi]$ into (3.3), we obtain $A^T_2 = J^T_{2k} A^T_2 J_{2n} = A^T_2$. It follows that $A^T_2 A_2 = A^T_2 = I_{2k}$, i.e., $A_2 \in V_{2k}(\mathbb{R}^{2n})$. For any $A_2 = [\Phi, \Psi; -\Psi, \Phi] \in V_{2k}(\mathbb{R}^{2n})$, (4.10) is satisfied. Thus, $A_2 \in Sp(2k, \mathbb{R}^{2n})$. Therefore, an equivalent definition of $M_2(2n, 2k)$ is given by

$$(4.17) \quad M_2(2n, 2k) = V_{2k}(\mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix} \mid \Phi, \Psi \in \mathbb{R}^{n \times k} \right\}. $$

**Lemma 4.3.** $V_k(\mathbb{C}^n)$ is isomorphic to $Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n})$.

**Proof.** Let $A_q = [\xi_1, \ldots, \xi_k], A_p = [\zeta_1, \ldots, \zeta_k], \text{ and } A_2 = [A_q, A_p] \in Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n})$. For each $i$, we have $||\xi_i|| = ||\zeta_i|| = 1$, and $\Omega(\xi_i, \zeta_i) = 1$. It follows that $||J_{2n}\xi_i|| = 1$, and the inner product $\langle J_{2n}\xi_i, \zeta_i \rangle = 1$. The Cauchy–Schwarz inequality states that $\langle J_{2n}\xi_i, \zeta_i \rangle = \langle J_{2n}\xi_i, \xi_i \rangle$, and two sides are equal iff $J_{2n}\xi_i$ and $\zeta_i$ are parallel. We must have that $J_{2n}\xi_i = \zeta_i$. Hence, $A$ must has the block form $A_2 = [A_q, J_{2n}A_q]$, or $[\Phi, -\Psi; \Phi, \Psi]$ if $A_q$ is written as $[\Phi; \Psi]$. Thus, $A_2 \in M_2(2n, 2k)$. By the proof in Lemma 4.2, we have $\Phi + i\Psi \in V_k(\mathbb{C}^n)$.

Conversely, if $\Phi + i\Psi \in V_k(\mathbb{C}^n)$, the mapping (1.14) yields that $\mathcal{A}(\Phi + i\Psi) \in Sp(2k, \mathbb{R}^{2n})$ and $\mathcal{A}(\Phi + i\Psi) \in V_{2k}(\mathbb{R}^{2n})$.

It is noted that $\mathcal{A}$ also preserves algebraic structures, as one can easily verify the following lemma.

**Lemma 4.4.** Let $C \in \mathbb{C}^{n_1 \times n_3}$ and $D \in \mathbb{C}^{n_2 \times n_3}$. Then, we have $\mathcal{A}(C)\mathcal{A}(D) = \mathcal{A}(CD)$, and $\mathcal{A}(C^H) = (\mathcal{A}(C))^T$.

**Theorem 4.5.** Suppose $Y \in \mathbb{R}^{2n \times N}$ is the snapshot matrix defined by (4.9). The symplectic matrix $A_2$ constructed by Algorithm 2 is the optimal solution that minimizes the least square error for representing $[Y, -J_{2n}Y]$ onto the column subspace within the manifold $M_2(2n, 2k)$.

**Proof.** By Lemma 4.4, the truncated SVD of $Z$ given by (4.13) yields

$$(4.18) \quad \mathcal{A}(Z) \approx \mathcal{A}(U)\mathcal{A}(\Sigma)\mathcal{A}(V^H) = \mathcal{A}(U)\mathcal{A}(\Sigma)(\mathcal{A}(V))^T. $$(

Since $U^H U = I_k$, By Lemma 4.4, we have

$$(\mathcal{A}(U))^T \mathcal{A}(U) = \mathcal{A}(U^H)\mathcal{A}(U) = \mathcal{A}(I_k) = I_{2k}. $$

Similarly, $(\mathcal{A}(V))^T \mathcal{A}(V) = I_{2k}$ holds due to $V^H V = I_k$. Moreover, $\mathcal{A}(\Sigma)$ is a real diagonal matrix that contains the first $2k$ dominant singular values of $\mathcal{A}(Z)$. Thus, (4.18) provides the truncated SVD for $\mathcal{A}(Z)$.

In Algorithm 2 the symplectic matrix is constructed by $A_2 = \mathcal{A}(U)$. Meanwhile, using the definition of $Z$ and $Y$, we have

$$\mathcal{A}(Z) = \begin{bmatrix} Y_1 & -Y_2 \\ Y_2 & Y_1 \end{bmatrix} = [Y, -J_{2n}Y]. $$

Thus, by the SVD method $A_2$ gives the basis such that the projection error of $[Y, -J_{2n}Y]$ onto the spanned subspace is minimal for a fixed dimension $2k$.

It should be emphasized that since the complex SVD method is designed to fit $[Y, -J_{2n}Y]$, rather than $Y$ itself, Algorithm 2 can only construct a near optimal matrix in $M_2$ to fit the empirical data.
So far, we have proposed three different algorithms to construct a symplectic basis matrix $A$. Corresponding to three manifolds with the inclusion maps:

$V_k(\mathbb{R}^n) \hookrightarrow V_k(\mathbb{C}^n) \overset{\text{def}}{\rightarrow} Sp(2k, \mathbb{R}^{2n})$,

we propose the cotangent lift, the complex SVD, and nonlinear optimization respectively to form a symplectic matrix, as listed in Table 4.1. The cotangent lift and the complex SVD are more efficient in offline computation, since direct SVD methods could be used to compute the basis matrix. However, as they are restricted to the block form, these methods search optimal basis matrices in submanifolds of $Sp(2k, \mathbb{R}^{2n})$, rather than $Sp(2k, \mathbb{R}^{2n})$ itself. Therefore, they sacrifice certain accuracy to fit empirical data in order to achieve stability guaranteed for reduced models and reduce the offline searching complexity. On the contrary, the nonlinear programming method requires more computational cost in the offline stage, since it computes several iterations to obtain a convergent value. However, it results in the best symplectic matrix that approximates a given snapshot ensemble.

5. Symplectic discrete empirical interpolation method (SDEIM). The complexity of the PSD with symplectic projection introduced above is approximately the same as the standard POD-Galerkin method for online computation. According to some previous studies, the standard Galerkin method generally cannot reduce the complexity for a large-scale system unless the vector field is a linear or quadratic function [7, 24]. Otherwise, the computation of the reduced system could be as expensive as, if not more than, the original system. Similarly, the PSD with symplectic projection cannot always construct a efficient reduced model when the original Hamiltonian system is nonlinear. Motivated by this fact, the SDEIM is proposed in this section to reduce the complexity of the nonlinear Hamiltonian system while simultaneously preserving the symplectic structure.

The original Hamiltonian, $X_H(y)$ can be split into a linear part and a nonlinear part, i.e., $H(y) = H_1(y) + H_2(y)$, such that $\nabla_y H_1(y) = Ly$ for a real symmetric matrix $L$, and $\nabla_y H_2(y) = F(y)$ for a nonlinear function $F$. Thus, the original system (2.2) can be rewritten as

\begin{equation}
\dot{y} = Ky + J_{2n}F(y),
\end{equation}

where $K = J_{2n}L \in \text{sp}(\mathbb{R}^{2n})$. Similar to (3.14), the reduced Hamiltonian system can be written as

\begin{equation}
\dot{x} = A^+(Ky + J_{2n}F(y)) = \tilde{K}x + J_{2k}A^T F(Ax).
\end{equation}

Unless $A^T F(Ax)$ can be analytically simplified, the computational complexity of (5.2) still depends on $2n$. In the worst scenario, one needs to first compute the matrix multiplication $y = Ax$, then compute the nonlinear vector term $F(y)$ in $\mathbb{R}^{2n}$, and
then project it back to $\mathbb{R}^{2k}$ by computing $A^T F$. As a result, the cost of the reduced PSD system is as expensive as, if not more than, the full system.

As an approximation of the symplectic Galerkin projection, the SDEIM is developed in this section that combine the PSD with the classical DEIM. Assume $F$ is a smooth function of $y$. If $F(y)$ resides approximately on the range of a $2n \times m$ matrix $V$, there exists a corresponding coefficient vector $\tau \in \mathbb{R}^m$ such that $F(y) \approx V \tau(y)$. The coefficient vector $\tau(y)$ can be determined by matching the nonlinear vector term at selected $m$ elements. Define a $2n \times m$ projection matrix

$$P := [e_{\beta_1}, \ldots, e_{\beta_m}],$$

where $e_{\beta_i}$ is the $\beta_i$th column of the identity matrix $I_{2n}$. Let $\beta = [\beta_1; \ldots; \beta_m] \in \mathbb{R}^m$ be an index vector. Then, left multiplication of $F(y)$ with $P^T$ projects a state variable or a vector field onto $m$ elements corresponding to the index vector $\beta$. Thus, we have $P^T F(y) = (P^T V) \tau(y)$. Suppose $P^T F(y)$ is nonsingular. Then, the coefficient vector $\tau(y)$ is computed from $\tau(y) = (P^T V)^{-1} P^T F(y)$. Thus, the approximation $\hat{F}(y)$ of the nonlinear vector term $F(y)$ becomes

$$\hat{F}(y) = V \tau(y) = V (P^T V)^{-1} P^T F(y),$$

and the reduced system from (5.1) can be approximated as

$$\dot{x} = \hat{K} x + J_{2k} A^T V (P^T V)^{-1} P^T F(Ax).$$

The standard DEIM can be used to determine the matrix $V$ in the above equation. However, a reduced POD-DEIM system is not necessarily to be Hamiltonian. To see this, define a $2k \times 2k$ matrix, $W = A^T V (P^T V)^{-1} P^T (A^+)^T$, and rewrite (5.3) in the following form

$$\dot{x} = \hat{K} x + J_{2k} W \nabla_x H_2(Ax).$$

Equation (5.6) is Hamiltonian iff $\hat{K} + J_{2n} W D^2 H_2(Ax) \in \mathfrak{sp}(\mathbb{R}^{2k})$, which implies that $WD^2 H_2(Ax)$ is a symmetric matrix for all $x$. Unfortunately, the standard DEIM does not always guarantee this condition.

One can form a nonlinear programming problem to search the optimal values of $A$, $V$, and $P$ to fit empirical data and satisfy the above condition. But this process could be very expensive for a large-scale problem. Thus, this section is devoted to discuss an easy-implemented algorithm to construct $A$, $V$, and $P$ with relatively low cost. Note that as long as

$$A = V \in \mathbb{M}_2(2n, 2k),$$

we have $A^T A = I_{2n}$ and $A^+ = A^T$. It follows that $W = I_{2k}$, and $WD^2 H_2(Ax)$ is symmetric. Thus, (5.7) is a sufficient condition for (5.6) to remain Hamiltonian. Replacing $V$ by $A$, and using $A^T A = I_{2k}$, we simplify the RHS of (5.5) and achieve the following definition.

**Definition 5.1.** The **SDEIM** of a nonlinear Hamilton’s equation $\dot{y} = J_{2n} \nabla_y H(y)$, or $\dot{y} = \hat{K} x + J_{2n} F(Ax)$, with an initial condition $y(0) = y_0$ is given by

$$\dot{x} = \tilde{K} x + J_{2k} (P^T A)^{-1} P^T F(Ax), \quad x_0 = A^T y_0,$$

where $A \in \mathbb{M}_2(2n, 2k)$, $\tilde{K} = A^T K A$, and $x_0$ is the initial condition of (5.8).
Since \( y(t) \in T^*Q \), \( [y(t); \nabla y H_2(y(t))], \) or \( [y(t); F(y(t))] \), can be considered as a trajectory of \( T^*(T^*Q) \). By assuming \( A = V \) in (5.7), we actually lift a mapping \( \phi : T^*R \rightarrow T^*Q \) to \( \phi_* : T(T^*R) \rightarrow T(T^*Q) \) via a \( 4n \times 4k \) matrix, \( \text{diag}(A, A) \). Using a similar idea from section 4.2 we can construct an extended snapshot ensemble in the phase space,

\[
\mathcal{X} := \{y(t_1), \ldots, y(t_N), F(y(t_1)), \ldots, F(y(t_N))\},
\]

that contains both snapshot \( y(t_i) \) and the nonlinear vector term \( F(y(t_i)) \). Both the cotangent lift method (in section 4.2) and the complex SVD method (in section 4.3) can be used to find a symplectic matrix \( A \in \mathbb{M}_2(2n, 2k) \) based on \( \mathcal{X} \).

**Algorithm 3** Greedy algorithm to construct an index vector of \( A \)

- **Require:** A symplectic matrix \( A = [a_1, \ldots, a_m] \in \mathbb{M}_2(2n, m) \).
- **Ensure:** Index vector \( \beta = [\beta_1; \ldots; \beta_m] \in \mathbb{R}^m \).

1. Select the first interpolation index \([\rho, \beta_1] = \max|\{a_1\}|. \) Initialize \( B = [a_1] \).
   
   for \( i = 2 \) to \( m \) do
   
   2. Solve the coefficient vector \( \tau \) to match \( a_i, B(\beta,:)\tau = a_i(\beta) \).
   
   3. Calculate the residual \( r = a_i - Br \).
   
   4. Select the interpolation index corresponding to the largest magnitude of the residual \( r, [\rho, \beta_1] = \max\{|r|\} \).
   
   5. Update \( B = [B, a_i], \beta = [\beta; \beta_i] \).

end for

A greedy algorithm in standard DEIM\(^3\) is applied here to construct the index vector \( \beta \), as listed in Algorithm 3. In step 4, \( [\rho, \beta_i] = \max\{|r|\} \) means \( \rho = |r(\beta_i)| = \max_{j=1,\ldots,2n}|r(j)| \). In step 5, we add a column vector \( a_i \) (and an element \( \beta_i \)) to a matrix \( B \) (and a vector \( \beta \)). Initially, we select the first interpolation index \( \beta_1 \in \{1, \ldots, n\} \) corresponding to the first basis function \( a_1 \) with largest magnitude. The remaining interpolation indices, \( \beta_i \) for \( i = 2, \ldots, m \), are respectively corresponding to the largest magnitude of the residual \( r \), where \( r \) is the residual or the error between the input basis \( a_i \) and its projection on a subspace spanned by \( B \). In \( \mathbb{M}_2 \), it has been proved that \( \rho \neq 0 \) implies that \( P^T A \) is nonsingular, which means that the approximation of \( F(y) \) in (5.4) is well-defined. Moreover, it is easy to verify that \( P^T F(y) = P^T F(y) \).

In terms of computational complexity of SDEIM, \( P \), \( A \) and \( J_{2k}(P^T A)^{-1} \) are calculated only once at the beginning. For each step in the online stage, the nonlinear vector term \( P^T F(Ax) \) is only evaluated on selected 2k elements of \( F(Ax) \). Thus, a significant speedup is obtained when \( k \ll n \), and \( m' \ll n \), where \( m' \) denotes the number of elements of \( Ax \) that is required to compute 2k elements of \( F(Ax) \).

Table 5.1 compares the standard POD-Galerkin method with the proposed symplectic model reduction method, and lists the main features and properties in the framework of symplectic model reduction. It serves as a short summary of section 5 through section 5.

6. Numerical Examples. In this section, the performance of symplectic model reduction is illustrated in numerical simulation of wave equations. After deriving the Hamiltonian form of general wave equations, we first study a linear wave equation

\(^3\)The MATLAB notations \( B(\beta,:) \) and \( a_i(\beta) \) are used here to represent the operation of selecting rows out of a matrix (or a vector).
Table 5.1
Complexity of the standard POD-Galerkin method and the Symplectic model reduction method

|                         | POD-Galerkin                                                                 | Symplectic model reduction                      |
|-------------------------|------------------------------------------------------------------------------|--------------------------------------------------|
| Original system         | \( y = f(y), \ y \in \mathbb{R}^n \)                                        | \( y = J_{2n} \nabla_y H(y), \ y \in \mathbb{R}^{2n} \) |
| Reduced state           | Orthogonal projection: \( x = \Phi^T y \in \mathbb{R}^k \)                  | Symplectic projection: \( x = A^T y \in \mathbb{R}^{2k} \) |
| Reduced system          | Galerkin projection: \( \dot{x} = \Phi^T f(\Phi x) \)                     | Symplectic Galerkin projection: \( \dot{x} = J_{2k} \nabla_x H(Ax) \) |
| Properties of reduced system | No stability guarantee                                                        | Energy preservation                                |
|                         |                                                                              | Volume preservation                                |
|                         |                                                                              | Stability preservation                              |
| Basis matrix            | Orthonormal: \( \Phi^T \Phi = I_k \)                                        | Symplectic: \( A^T J_{2n} A = J_{2k} \)         |
| Domain of basis matrix  | Stiefel manifold \( V_k(\mathbb{R}^n) \)                                   | Symplectic Stiefel manifold \( Sp(2k, \mathbb{R}^{2n}) \) |
| Constructing basis matrix | Proper orthogonal decomposition (POD)                                 | Proper symplectic decomposition (PSD):           |
|                         |                                                                              | (a) Nonlinear programming                          |
|                         |                                                                              | (b) Cotangent lift                                 |
|                         |                                                                              | (c) Complex SVD                                   |
| Simplifying nonlinear terms | Standard DEIM                                                                 | SDEIM                                             |

numerically and focus on demonstrating the capability of symplectic model reduction algorithms to deliver stability-preserving reduced models. Then we simulate the nonlinear sine-Gordon equation to illustrate that the SDEIM is able to deliver accurate and long-time stable results with significant speedups.

### 6.1. Hamiltonian Formulation for Wave Equation

Let \( u = u(t, x) \). Consider the one-dimensional semi-linear wave equation with constant moving speed \( c \) and a nonlinear vector term \( g(u) \),

\[
\dddot{u} = c^2 u_{xx} - g(u),
\]

on space \( x \in [0, L] \). With the generalized coordinates \( q = u \) and the generalized momenta \( p = \dot{u} \), the Hamiltonian PDE associated with (6.1) is given by

\[
\dot{q} = \frac{\delta H}{\delta p}, \quad \dot{p} = -\frac{\delta H}{\delta q},
\]

where the Hamiltonian is defined as

\[
H(q, p) = \int_0^L dx \left[ \frac{1}{2} p^2 + \frac{1}{2} c^2 q_x^2 + G(u) \right], \quad G'(u) = g(u).
\]

For Hamiltonian PDEs, both symplectic and multisymplectic schemes [5] can be applied to construct a fully resolved model that preserve the symplectic structure of the original PDE. In this paper, we choose a symplectic scheme with spatial semi-discretization of \( n \) equally spaced grid points. The spatial discretized Hamiltonian is given by

\[
\mathcal{H}_{\Delta x}(y) = \sum_{i=1}^n \Delta x \left[ \frac{1}{2} p_i^2 + \frac{c^2(q_{i+1} - q_i)^2}{4\Delta x^2} + \frac{c^2(q_i - q_{i-1})^2}{4\Delta x^2} + G(q_i) \right],
\]
where \( q_i := u(t,x_i), p_i := \dot{u}(t,x_i) \), \( y := [q_1; \ldots; q_n; p_1; \ldots; p_n] \in \mathbb{R}^{2n}, x_i = i\Delta x, \) \( i \in \{1, \ldots, n\} \) are the points of a uniform discretization in space. In the limit \( \Delta x \rightarrow 0 \) and \( n\Delta x = L, \) (6.3) converges to (6.5). Moreover, the periodic boundary condition requires that \( q_0 = q_n \) and \( q_{n+1} = q_1 \); the Dirichlet boundary condition gives specified values for \( q_0 \) and \( q_{n+1} \); and the Neumann boundary condition requires that \( q_0 = q_1 \) and \( q_{n+1} = q_n \). Now, we have a Hamiltonian system of coupled nonlinear ODEs,

\[
\dot{y} = J\Delta x \nabla_y \mathcal{H}_{\Delta x}, \quad J_{\Delta x} = \frac{J_{2n}}{\Delta x}.
\]

Let \( D_{xx} \in \mathbb{R}^{n \times n} \) denote the central difference approximation of the second order spatial derivative, and

\[
D_{xx} := \frac{1}{\Delta x^2} \begin{bmatrix} a & 1 & b \\ 1 & -2 & 1 \\ & \ddots & \ddots & \ddots \\ b & 1 & a \end{bmatrix}.
\]

For the periodic boundary condition, \( b = 1, a = -2 \); for the Dirichlet boundary condition, \( b = 0, a = -2 \); and for the Neumann boundary condition, \( b = 0, a = -1 \). Thus, (6.5) can be written in the form \( \dot{y} = Ky + y_{bd} + J_{2n}F(y) \), where the Hamiltonian matrix is given by

\[
K = \begin{bmatrix} 0_n & I_n \\ c^2D_{xx} & 0_n \end{bmatrix},
\]

and the nonlinear function \( F(y) \) is a vector in \( \mathbb{R}^{2n} \) with zeros in the last \( n \) elements, i.e., \( F(y) = [g(q); 0_{n \times 1}] \). If the Dirichlet boundary condition is used, \( y_{bd} = [0_{n \times 1}; q_0, 0_{(n-2) \times 1}; q_{n+1}] \in \mathbb{R}^{2n}, \) otherwise \( y_{bd} = 0_{2n \times 1} \). Time discretization can be achieved by using the implicit symplectic integrator scheme (2.3). If \( G(u) = 0, \) the successive over relaxation can be used to update the linear system for each time step; otherwise, the system is nonlinear and the Newton iteration can be used to time advance one step.

### 6.2. Linear Wave Equation

For our numerical experiments, we first study a linear system with \( G(u) = g(u) = 0 \) and with the periodic boundary condition on the interval \([0, L]\). Let \( s = 10 \times |x - \frac{L}{2}| \), and let \( h(x) \) be a cubic spline function, which is \( 1 - \frac{3}{2}s^2 + \frac{3}{4}s^3 \) if \( 0 \leq s \leq 1 \), \( \frac{1}{4}(2-s)^3 \) if \( 1 < s \leq 2 \), and 0 if \( s > 2 \). The initial condition is provided by

\[
y_0 = [h(x_1); \ldots; h(x_n); 0_{n \times 1}],
\]

which gives rise to a periodic system with wave propagating in both directions of \( x \) in a periodic domain. The full model (reference benchmark solver) is computed using the following parameter set:

| Parameter                       | Value          |
|---------------------------------|----------------|
| Space interval                  | \( L = 1 \)    |
| Number of grid points           | \( n = 500 \)   |
| Space discretization step       | \( \Delta x = L/n = 0.002 \) |
| Time interval                   | \( T = 50 \)    |
| Time discretization step        | \( \delta t = 0.01 \) |
| Speed of the wave               | \( c = 0.1 \)   |
The solution profile can be seen in Figure 6.1(a), where the initial state and the wave amplitude at \( t = 2.5 \) and \( t = 5 \) are shown. The snapshot ensemble takes 101 snapshots from the benchmark solution trajectory with uniform interval (\( \Delta t = 0.5 \)). We first compare the standard POD-Galerkin method and the cotangent lift method with the symplectic projection method. For short-time integration, both methods could obtain very accurate results by taking the first 20 modes. In Figure 6.1(b), the blue line represents the singular values of the snapshot matrix \( Y \) for the standard POD method. Suppose \( \{ \lambda_1, \ldots, \lambda_k \} \) denotes the singular values of the extended snapshot matrix \( X \) in the cotangent lift method. The red line represents the duplicated singular values \( \{ \lambda_1, \lambda_1, \ldots, \lambda_k, \lambda_k \} \) of \( X \) corresponding to the symplectic basis matrix \( A_1 = \text{diag}(\Phi, \Phi) \). The black line represents the duplicated singular values of the complex snapshot matrix \( Z \) in the complex SVD method. This plot demonstrates that when we use the cotangent lift or the complex SVD, we do sacrifice certain accuracy to fit empirical data in the exchange of stability preservation and energy preservation.

![Figure 6.1](image)

**Figure 6.1.** (Color online.) (a) The solution \( u(t, x) \) at \( t = 0 \), \( t = 2.5 \) and \( t = 5 \) of the linear wave equation (6.1) with \( g(u) = 0 \) and the initial condition (6.8) for \( t \in [0, 50] \). We compare the standard POD-Galerkin method (referred by “POD” in the legend) and the cotangent lift method with the symplectic projection (referred by “symplectic” in the legend). For short-time integration, both methods could obtain very accurate results by taking the first 20 modes. (b) Plot the singular values of the POD method, the cotangent lift method, and the complex SVD method.

Using more modes, one may expect both methods to produce more accurate solutions. However, as Figure 6.2 indicates, the standard POD-Galerkin method blows up quickly when we use more than 30 POD modes. Moreover, the more modes we use, the earlier the system blows up. This phenomena implies that the POD-Galerkin projection may be unstable for some hyperbolic equations, and can generate unpredictable results. While reduced systems constructed by symplectic methods show some numerical error, this error could be systematically reduced by using more modes. Figure 6.3 indicates that the \( L_2 \) error of the cotangent lift and the complex SVD converges to a small number when more modes are used, while the \( L_2 \) norm error of the POD method is small only for \( k = 20 \) for the cases tested with \( k = 10, 20, \ldots, 80 \).

Using 20 modes (\( k = 20 \)), both POD-Galerkin method and the symplectic methods can generate reduced models with relatively small error. However, as Figure 6.4 demonstrates, symplectic methods preserve the system energy \( E \), no matter how many modes are used in constructing the reduced system. However, for \( k = 20 \), the POD-Galerkin method decreases the energy in the time domain \([0, 50]\). For \( k = 30 \) or 40, the energy quickly grows and indicates an unstable reduced model.
SYMPLECTIC MODEL REDUCTION OF HAMILTONIAN SYSTEMS

Figure 6.2. (Color online.) (a) The evolution of instant $L_2$ error, $\|e(t)\| := \|\hat{u}(t) - u(t)\|$, between the benchmark solution $u(t)$ and approximating solutions $\hat{u}(t)$ of the linear wave equation. We select different subspace dimension, $k$, for different model reduction techniques. (b) The instant $L_2$ error, $\|e(t)\|$, for $t \in [0,0.5]$.

Figure 6.3. (Color online.) The $L_2$ norm of $\|e(t)\|$ for the linear wave equation with different subspace dimension $k$. For the standard POD-Galerkin method, we only compute the error for $k = 10$ and $k = 20$. When $k > 20$, the reduced system blows up in the interested time domain $[0, 50]$ and $\|e(t)\|_2$ becomes infinite.

Figure 6.4. (Color online.) (a) The evolution of the system energy $E(t)$ of the linear wave equation, which is the same as the Hamiltonian $H_{\Delta_t}(y(t))$ defined in (6.4) with $G(u) = 0$, of reduced systems for $t \in [0,50]$. (b) The system energy $E(t)$ of reduced systems for $t \in [0,0.5]$. Initially, $E(0) = 0.075$. 
Stability Preservation of Symplectic Model Reduction. To explain our observations mentioned above, we study the stability of the linear wave equation. According to [4], the eigenvalues \( \gamma_i \) \((i = 1, \ldots, n)\) of \( D_{xx} \) (6.6) with periodic boundary conditions are given by

\[
\beta_i = -\frac{2}{\Delta x^2} \left[ 1 - \cos \left( \frac{2\pi i}{n} \right) \right],
\]

and the corresponding eigenvectors are given by

\[
w_i = \frac{1}{\sqrt{n}} \left[ e^{-2\pi i/n}, \ldots, e^{-2\pi i(n-1)/n}, 1 \right].
\]

It follows that the eigenvalues of the Hamiltonian matrix \( K \) in (6.7) are given by \( 2n \) pure imaginary numbers \( \pm \{\iota \gamma_i\}_{i=1}^n \), where \( \gamma_i = c\sqrt{-\beta_i} \) for \( k = 1, \ldots, n \); and the corresponding eigenvectors are given by

\[
\xi_i = \frac{1}{\sqrt{1 + w_i^2}} \left[ w_i, \iota \gamma_i w_i \right]; \quad \zeta_i = \frac{1}{\sqrt{1 + w_i^2}} \left[ w_i, -\iota \gamma_i w_i \right].
\]

Since \( \xi_n = \zeta_n = \frac{1}{\sqrt{n}} \left[ 1, 0, \ldots, 0 \right] \) by the above equation, we can redefine \( \zeta_n \) to be \( \zeta_n = \frac{1}{\sqrt{n}} \left[ 0, 1, \ldots, 1 \right] \). Thus, we can choose an insertable matrix \( Q = [\xi_1, \zeta_1, \ldots, \xi_n, \zeta_n] \) so that \( K \) is transformed to a real Jordan form

\[
Q^{-1}KQ = \text{diag} \left\{ \iota \gamma_1, -\iota \gamma_1, \ldots, \iota \gamma_{n-1}, -\iota \gamma_{n-1}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\}.
\]

Although \( K \) contains an unstable mode \( \zeta_n \), the projection coefficient of initial condition (6.8) onto this mode vanishes, i.e., \( c_n y_0 = 0 \). Thus, the original system is bounded for all \( t \).

Next, we consider the reduced model constructed by the symplectic projection. By (6.3), the system energy \( H(y) \) of the linear wave equation with the periodic boundary condition can be expressed as

\[
H(y) = \frac{\Delta x}{2} \sum_{i=1}^{n} p_i^2 + \frac{c^2}{2\Delta x} \sum_{i=1}^{n} (q_i - q_{i-1})^2,
\]

where \( q_0 = q_n \). Then \( H(y) \geq 0 \), and the equality holds if and only if \( y \) is parallel to \( \xi_n \). Let \( S^{2n-1} \) denote the unit sphere from the origin of \( \mathbb{R}^{2n} \), and let \( S_0 = S^{2n-1} \cap \text{Range}(A) \). If \( \xi_n \notin \text{Range}(A) \), we have \( H(z) \geq 0 \) for all \( z \in S_0 \). Since \( S_0 \) is closed in \( \mathbb{R}^{2n} \) and \( H(z) \) is a continuous function of \( z \), there exist a positive constant \( \varepsilon \) such that \( H(z) \geq \varepsilon \) for all \( z \in S_0 \). Let \( y \in \text{Range}(A) \), and \( z = y/\| y \| \). Since \( H(y) \) is a homogeneous function of \( y \), we have \( H(y) = \| y \|^2 H(z) \geq \varepsilon \| y \|^2 \) if \( \xi_n \notin \text{Range}(A) \).

In our numerical simulations, we do observe that \( AA^T \xi_n \) is nonzero for both the cotangent lift and the complex SVD. Since the reduced system is Hamiltonian and evolves in \( \text{Range}(A) \), the energy is conserved during time evolution. Therefore, the reduced system constructed by the symplectic projection are bounded for all \( t \), i.e., the symplectic projection preserves the stability of the linear wave equation (6.1) with \( f = 0 \) and initial condition (6.8).

Instability of POD-Galerkin. Since the standard POD-Galerkin method does not preserve the system energy, there are no mechanisms similar to our symplectic approach that limit the solution trajectory in a bounded region. Therefore, it is entirely
possible that the reduced system may blow up. To corroborate this claim, let $\lambda_*$ denote the eigenvalue of $\Phi^T K \Phi$ with maximal real part and $\xi_*$ denote a corresponding eigenvector with unit length. Additionally, let $a_* = \xi^T y_0$ denote the projection coefficient of $y_0$ onto $\xi_*$. The following table indicates that for different subspace dimensions $k$, we have $\text{Re}(\lambda_*) > 0$ and $a_* \neq 0$. Since the solution has an exponential term $a_* \exp(\lambda_* t) \xi_*$, the reduced system constructed by the Galerkin projection is always unstable for long-time integration.

Assume that $\Phi_k$ and $\Phi_l$ respectively contain the first $k$ and $l$ dominant modes. If $k < l$, then $\Phi_k^T K \Phi_k$ is a submatrix of $\Phi_l^T K \Phi_l$, and $\|\Phi_k^T K \Phi_k\| < \|\Phi_l^T K \Phi_l\|$ holds. As $\text{Re}(\lambda_*) \leq |\lambda_*| \leq \|\Phi^T K \Phi\|$, the matrix norm of $\Phi^T K \Phi$ provides an upper bound for $\text{Re}(\lambda_*)$. Thus, the upper bound of $\text{Re}(\lambda_*)$ is a monotonically increasing function of $k$.

The following table also shows that $\text{Re}(\lambda_*)$ with 80 modes is much larger than $\text{Re}(\lambda_*)$ with 20 modes. This explains why a reduced POD system with 80 modes blows up faster than a reduced POD system with 20 modes. Although for $k = 20$, POD can produce a reduced model with reasonable accuracy for a short time domain $[0, 50]$, we can still observe that for a large enough integration time, say $t > 10^4$, the reduced POD system will blow up exponentially.

| $k$   | 10   | 20   | 30   | 40   | 50   | 60   | 70   | 80   |
|-------|------|------|------|------|------|------|------|------|
| $\lambda_*$ | 2.83e-3±5.98e-3 | 0.0446 | 15.3 | 19.8 | 16.6 | 17.9 | 19.8±1.9 | 111  |
| $a_*$  | -0.435±0.915i | 4.60e-3 | 6.90e-5 | -1.44e-4 | 1.66e-4 | -1.41e-4 | 1.01e-5±2.88e-5 | -8.97e-6 |

### 6.3. Sine-Gordon Equation.

Next, we consider a special nonlinear wave equation with $G(u) = 1 - \cos(u)$, $g(u) = \sin(u)$ and $c = 1$, which corresponds to the sine-Gordon equation. This equation was first studied in the 1970s, and then appears in a number of physical applications, including applications in relativistic field theory, Josephson junctions or mechanical transmission lines \[28\]. One can show that the sine-Gordon equation admits a localized solitary wave solution,

$$u(t, x) = 4 \arctan \left( \exp \left( \pm \frac{x - x_0 - vt}{\sqrt{1 - v^2}} \right) \right),$$

which travels with the speed $|v| < 1$. The $\pm$ signs correspond to localized solutions which are called “kink” and “antikink”, respectively \[28\].

In our simulations, the full model is solved for the “kink” case on the interval $[0, L]$ with Dirichlet boundary conditions ($u(t, 0) = 0, u(t, 1) = 2\pi$) using the following parameter set:

| Parameter                      | Value       |
|--------------------------------|-------------|
| space interval                 | $L = 50$    |
| Number of grid points          | $n = 2000$  |
| Space discretization step      | $\Delta x = L/n = 0.025$ |
| Time interval                  | $T = 150$   |
| Time discretization step       | $\delta t = 0.0125$ |
| Speed of the wave              | $v = 0.2$   |

For the SDEIM, the cotangent lift method is used to construct a symplectic basis matrix $A_1 = \text{diag}(\Phi, \Phi)$. Specifically, after collecting a snapshot ensemble \[5.9\] in the phase space, we construct an extended snapshot matrix

$$[q(t_1), \ldots, q(t_N), p(t_1), \ldots, p(t_N), f(q(t_1)), \ldots, f(q(t_N))].$$

The standard SVD is then applied to construct $\Phi$. 

LIQIAN PENG AND KAMRAN MOHSENI

Figure 6.5. (Color online.) (a) The field $u(t, x)$ at $t = 0$, $t = 25$ and $t = 75$ of the sine-Gordon equation (6.1) with $g(u) = \sin(u)$ and the “kink” solution (6.9). We compare the standard POD-Galerkin method (referred to “POD” in the legend) and the cotangent lift method with the symplectic projection (referred to “symplectic” in the legend). For short-time integration, both methods could obtain very accurate results by taking the first 60 modes. (b) Plot the singular values of different model reduction techniques.

The “kink” solution profile can be seen from Figure 6.5(a), where the initial state and the state at $t = 25$ and $t = 75$ are shown. The snapshot ensemble takes 1201 snapshots from the solution trajectory, solved by the full model with uniform interval $(\Delta t = 0.125)$. We first compare the standard POD-Galerkin method and the cotangent lift method with the symplectic projection. If the first 60 modes are used to construct reduced systems, accurate results can be obtained by both methods. In Figure 6.5(b), we plot the (duplicated) singular values of snapshot matrices by different methods, which demonstrates that POD is better to fit empirical state vectors than the cotangent lift and the complex SVD while standard DEIM is better to fit empirical nonlinear vector terms than the SDEIM.

Figure 6.6 illustrates that, during time evolution, all symplectic schemes (including the cotangent lift method, the complex SVD method, and the symplectic DEIM) yield low computational errors with appropriate subspace dimension, $k$, while nonsymplectic schemes (including the standard POD method and the standard DEIM) yield large numerical error or even blow up. In Figure 6.7, all symplectic schemes can
SYMPLECTIC MODEL REDUCTION OF HAMILTONIAN SYSTEMS

Figure 6.7. (Color online.) (a) The evolution of the system energy $E(t)$, which is the same as the Hamiltonian $H_{\Delta}(y(t))$ defined in (6.4) with $G(u) = 1 - \cos(u)$, of the sine-Gordon equation for $t \in [0, 150]$. (b) The system energy $E(t)$ for $t \in [0, 5]$. Initially, $E(0) = 4.25$.

Figure 6.8. (Color online.) (a) Compare the $L_2$ norm of $\|e(t)\|$ of different model reduction techniques with different subspace dimension, $k$, for the sine-Gordon equation. For the standard POD-Galerkin method, we only compute the error for $k = 40$ and $k = 60$. When $k > 60$, the reduced POD system blows up in the interested time domain $[0, 150]$ and $\|e(t)\|_2$ becomes infinite. (b) Compare the running time of different model reduction techniques.

effectively preserve the system energy. In contrast, the standard POD method results in infinite energy while the standard DEIM yields large and unstable energy. Figure 6.8(a) indicates that by using more modes all symplectic schemes could obtain better accuracy and approaches to the full model. However, the POD method yields infinite solution for $k > 60$ for the interested time domain, and the standard DEIM maintains large numerical error for all $k$.

By the analysis in Section 5, we know that a direct use of the PSD with the symplectic projection is not able to obtain any speedups for the sine–Gordon equation, since it contains a nonlinear vector term. Numerical results in Figure 6.8(b) also verify this point. Especially, the running time for the cotangent lift and the complex SVD with the symplectic projection is even larger than the running time for the full mode. On the other hand, the SDEIM approximation with the cotangent lift could significantly improve the efficiency and reduce the running time.

The stability of symplectic reduced models could be derived by its energy conservation property. By (6.3), the system energy $H(y)$ of the sine–Gordon equation with
the Dirichlet boundary condition is given by

$$H(y) = \frac{\Delta x}{2} \sum_{i=1}^{n} p_i^2 + \Delta x \sum_{i=1}^{n} |1 - \cos(q_i)| + \frac{q_1^2}{4\Delta x} + \frac{1}{2\Delta x} \sum_{i=2}^{n} (q_i - q_{i-1})^2 + \frac{(q_n - 2\pi)^2}{4\Delta x}.$$ 

If \(y(t)\) denote the solution trajectory, we have \(H(y(t)) = E\) for a constant \(E\). Since each term in the above equation is nonnegative, we must have \(|p_i| \leq \sqrt{2E/\Delta x}\), \(|q_1| \leq 2\sqrt{E/\Delta x}\), and \(|q_i| \leq |q_{i-1}| + \sqrt{2E/\Delta x}\) for \(i \geq 2\). In other words, there exists a positive number \(M\), for any state \(z \in \mathbb{R}^{2n}\), as long as \(\|z\| > M\), we have \(H(z) > E\). Therefore, by Theorem 3.7 both the original system and reduced systems constructed by the symplectic projection are bounded for all \(t\).

7. Conclusion. In this paper, a symplectic model reduction technique was proposed for the reduced-order modeling of large-scale Hamiltonian systems. We first define the symplectic projection that yields reduced-order systems that also remain Hamiltonian. Several proper symplectic decomposition (PSD) algorithms, such as the nonlinear programming method, the cotangent lift method, and the complex SVD method were developed to generate a symplectic matrix that spans a low-dimensional symplectic subspace. Because the symplectic model reduction preserves the symplectic structure, it also preserves the system energy and the volume of flow. Moreover, preserving the symplectic structure could guarantee long-time stability of the reduced system, which is not preserved in the standard POD-Galerkin approach. Thus, the proposed technique is suitable for long-time integration, especially when the original PDEs are hyperbolic and do not have any natural dissipative mechanism to stabilize them. Since the symplectic projection can only speed up linear or quadratic problems, the PSD was also combined with DEIM that can effectively reduce the complexity of the nonlinear vector term. Because the complexity of the SDEIM does not depend on the dimension of the original system, a significant speedup can be obtained for a general nonlinear problem. The capability of the symplectic model reduction to solve a large-scale system with high accuracy, good efficiency, and stability preservation were demonstrated for linear and nonlinear wave equations.

REFERENCES

[1] M. Amabili, A. Sarkar, and M. P. Paidoussis, Reduced-order models for nonlinear vibrations of cylindrical shells via the proper orthogonal decomposition method, J. Fluids Struct., 18 (2003), pp. 227–250.
[2] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, A survey of model reduction methods for large-scale systems, Contemp. Math., 280 (2001), pp. 193–219.
[3] P. Astrid, S. Weiland, K. Willcox, and T. Backx, Missing point estimation in models described by proper orthogonal decomposition, IEEE Trans. Automat. Control, 53 (2008), pp. 2237–2251.
[4] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, An ‘empirical interpolation’ method: Application to efficient reduced-basis discretization of partial differential equations, C. R. Acad. Sci. Paris, Ser. I, 339 (2004), pp. 667–672.
[5] T. J. Bridges and S. Reich, Numerical method for Hamiltonian PDEs, J. Phys. A: Math. Gen., 39 (2006), pp. 5287–5320.
[6] K. Carlberg, R. Tuminaro, and P. Boggs, Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics, arXiv:1401.8044, (2014).
[7] S. Chaturantabut and D. C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. Sci. Comput., 32 (2010), pp. 2737–2764.
[8] Y. Chen and J. White, A quadratic method for nonlinear model order reduction, in Proc. Int. Conf. Modeling and Simulation of Microsystems, 2000, pp. 477–480.
[9] M. Drohmann, B. Haasdonk, and M. Ohlberger, Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation, SIAM J. Sci. Comput., 34 (2012), pp. A937–A969.

[10] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations, M2AN Math. Model. Numer. Anal., 41 (2007), pp. 575–605.

[11] E. Hairer, C. Lubich, and G. Wanner, Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations, Springer, New York, 2006.

[12] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge Univ. Press, Cambridge, UK, 2nd ed., 2002.

[13] S. Lall, P. Kryst, and J. E. Marsden, Structure-preserving model reduction for mechanical systems, Physica D, 184 (2003), pp. 304–318.

[14] R. J. LeVeque, Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems, Classics in Applied Mathematics, SIAM, 2007.

[15] J. E. Marsden and T. Ratiu, Introduction to Mechanics and Symmetry, vol. 17 of Texts in Applied Mathematics, Springer, 2nd ed., 2003.

[16] R. I. McLachlan and G. R. W. Quispel, Geometric integrators for ODEs, J. Phys. A: Math. Gen., 39 (2006), pp. 5251–5285.

[17] J. D. Meiss, Differential Dynamical Systems, SIAM, Philadelphia, PA, 2007.

[18] K. R. Meyer, G. R. Hall, and D. Offin, Introduction to Hamiltonian Dynamical Systems and the N-Body Problem, vol. 90 of Applied Mathematical Sciences, Springer, 2nd ed., 2009.

[19] B. C. Moore, Principal component analysis in linear systems: Controllability, observability, and model reduction, IEEE Trans. Automat. Contr., 26 (1981), pp. 17–32.

[20] P. A. Parrilo, S. Lall, F. Paganini, G. C. Verghese, B. C. Leiseste, and J. E. Marsden, Model reduction for analysis of cascading failures in power systems, in Proc. American Control Conference, vol. 6, San Diego, CA, June 1999, pp. 4208–4212.

[21] L. Peng and K. Mohseni, An online manifold learning approach for model reduction of dynamical systems, in press SIAM J. Numer. Anal., (2014).

[22] T. A. Porsching, Estimation of the error in the reduced basis method solution of nonlinear equations, Mathematics of Computation, 45 (1985), pp. 487–496.

[23] S. Prajna, POD model reduction with stability guarantee, in Proc. 42nd IEEE Conf. Decision Control, Dec. 2003.

[24] M. Rathinam and L. R. Petzold, A new look at proper orthogonal decomposition, SIAM J. Numer. Anal., 41 (2003), pp. 1893–1925.

[25] M. Rewienski and J. White, A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices, IEEE Trans. Comput. Aided Des. Integr. Circuits Syst., 22 (2003), pp. 155–170.

[26] C. W. Rowley, T. Colonius, and R. M. Murra, Model reduction for compressible flows using POD and Galerkin projection, Physica D, 189 (2004), pp. 115–129.

[27] T. Bu Thanh An D M. Damodaran and K. Willcox, Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition, AIAA Journal, 42 (2004), pp. 1505–1516.

[28] G. B. Whitham, Linear and Nonlinear Waves, Wiley-Interscience, 1999.

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