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 save the computational cost for the simplification of large-scale Hamiltonian systems while preserving the symplectic structure. As an analogy to the classical proper orthogonal decomposition (POD)-Galerkin approach, PSD is designed to build a symplectic subspace to fit empirical data, while the symplectic Galerkin projection constructs a reduced Hamiltonian system on the symplectic subspace. For practical use, we introduce three algorithms for PSD, which are based upon the cotangent lift, complex singular value decomposition, and nonlinear programming. The proposed technique has been proven to preserve system energy and stability. Moreover, PSD can be combined with the discrete empirical interpolation method to reduce the computational cost for nonlinear Hamiltonian systems. Owing to these properties, the proposed technique is better suited than the classical POD-Galerkin approach for model reduction of Hamiltonian systems, especially when long-time integration is required. 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 systems, proper symplectic decomposition, symplectic Galerkin projection, symplectic structure preservation, energy preservation, stability preservation, symplectic discrete empirical interpolation method

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

DOI. 10.1137/140978922

1. Introduction. To save computational costs, model reduction seeks to approximate high-dimensional dynamical systems using simpler, lower-order ones that can capture the dominant dynamic properties. The need for model reduction arises because, in many cases, direct numerical simulations are often so computationally intensive that they either cannot be performed as often as needed or are performed only in special circumstances. See [2] for a survey on the available model reduction techniques.

Among these techniques, the proper orthogonal decomposition (POD) with Galerkin projection, which was first introduced by Moore [24], has wide applications in many fields of science and engineering, such as electric circuit analysis [25], structural dynamics [1], and fluid mechanics [17, 29], to list a few. As an empirical model reduction technique, the POD-Galerkin approach (or POD for short) involves an offline-online splitting methodology. In the offline stage, empirical data is generated by direct numerical simulations of the original system. If the original system is represented by a PDE, a discretized high-dimensional model can be derived by the finite difference, finite element, and finite volume methods. The POD can be applied to compute an optimal subspace to fit the empirical data. A reduced system is then

---

†Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250 (liqianpeng@ufl.edu).

‡Department of Mechanical and Aerospace Engineering, and Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6250 (mohseni@ufl.edu).

constructed by projecting the high-dimensional system to this subspace. In the online stage, one can solve the reduced system in the low-dimensional subspace. Recently, many variants of POD-Galerkin have been developed to reduce the complexity of evaluating the nonlinear term of the vector field, such as trajectory piecewise linear approximation [28], missing point estimation [3], gappy POD [6, 32, 7], empirical interpolation [4, 13], and the discrete empirical interpolation method (DEIM) [9, 10]. Thanks to these methods, the computational complexity during the online stage is independent of the dimension of the high-dimensional model.

Generally, the classical POD method is not guaranteed to yield a stable reduced system, even if the original system is stable [27, 26]. The instability of a reduced system is often accompanied by blowup of system energy and 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 lagrangian systems, Lall, Krysl, and Marsden showed that performing a Galerkin projection on the configuration space and lifting the projection to the phase space lead to reduced systems that preserve the original Lagrangian structure [18]. In order to reduce the complexity of nonlinear Lagrangian systems, Carlberg, Tuminaro, and Boggs combined Lall's method with the gappy POD to derive reduced nonlinear Lagrangian systems [8]. In the control community, the balanced truncation [16], moment matching [30], and tangential interpolation [14] approaches were used to preserve the port-Hamiltonian structure.

In this paper, we propose a new model reduction technique, proper symplectic decomposition (PSD), that preserves the symplectic structure underlying the Hamiltonian mechanics. Our main focus is to develop a basic framework behind symplectic model reduction, which allows us to derive energy preservation and stability preservation. The proposed technique yields reduced Hamiltonian systems which are applicable to long-time integration. Compared with other empirical model reduction algorithms that preserve system energy, the PSD is easier for applications; the computation complexity can be the same magnitude as the original POD and DEIM for both offline and online stages. The PSD also increases the 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. Section 3 presents the symplectic projection, which constructs reduced Hamiltonian systems. In section 4, three different PSD algorithms are proposed to construct a symplectic matrix, including the cotangent lift, complex singular value decomposition (SVD), and nonlinear programming (NLP). 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, and 5 are respectively analogous to 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. Hamiltonian system and symplectic integrator. Let $\mathbb{V}$ be a vector space of dimension $2n$. A symplectic form on $\mathbb{V}$ is a nondegenerate and alternating bilinear form, $\Omega : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$. The pair $(\mathbb{V}, \Omega)$ is called a symplectic vector space. Assigning a symplectic form Ω to $\mathbb{V}$ is referred to as giving $\mathbb{V}$ a symplectic structure. With canonical coordinates on $\mathbb{V}$ denoted by $(q_1, \ldots, q_n, p_1, \ldots, p_n)$, Ω takes a canonical

form, $\Omega = \sum_{i=1}^{n} dq_i \wedge dp_i$. Throughout this paper, we implicitly assume that $\mathbb{V}$ is defined over the field $\mathbb{R}$, which means $\mathbb{V} \cong \mathbb{R}^{2n}$. Moreover, for all $v_1, v_2 \in \mathbb{V}$, Ω is represented by the Poisson matrix J_{2n}, i.e.,

$$\Omega(v_1, v_2) = v_1^T J_{2n} v_2, \qquad J_{2n} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix},$$

where 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 $J_{2n} J_{2n} = J_{2n}^T J_{2n}^T = -I_{2n}$, where the superscript T represents the transpose of a matrix.

Let $H : \mathbb{V} \to \mathbb{R}$ denote a smooth Hamiltonian function. The time evolution of an autonomous Hamiltonian system is given by

$$(2.1) \qquad \dot{q} = \nabla_p H(q, p), \qquad \dot{p} = -\nabla_q H(q, p),$$

where $q = [q_1; \ldots; q_n] \in \mathbb{R}^n$, and $p = [p_1; \ldots; p_n] \in \mathbb{R}^n$. We abstract this formulation by introducing the phase space variable $x = [q; p]$[1] and the abstract Hamiltonian differential equation

$$(2.2) \qquad \dot{x} = X_H(x),$$

where $X_H(x) := J_{2n} \nabla_x H(x)$ is the Hamiltonian vector field. The flow Ψ_t of X_H is a symplectomorphism, meaning that it conserves the symplectic form Ω, the system Hamiltonian H, and the volume of flow Θ [20].

Symplectic integrators are numerical schemes for solving a Hamiltonian system while preserving the underlying symplectic structure. If the symplectic structure is preserved, then the flow volume and system energy are automatically conserved during time integration. By virtue of these advantages, symplectic integrators have been widely applied to long-time integrations of molecular dynamics, discrete element methods, accelerator physics, and celestial mechanics [15].

Let δt denote the unit step for time integration. The symplectic Euler methods

$$\begin{cases} q^{j+1} = q^j + \delta t \nabla_p H(q^{j+1}, p^j) \\ p^{j+1} = p^j - \delta t \nabla_q H(q^{j+1}, p^j) \end{cases} \quad \text{or} \quad \begin{cases} q^{j+1} = q^j + \delta t \nabla_p H(q^j, p^{j+1}) \\ p^{j+1} = p^j - \delta t \nabla_q H(q^j, p^{j+1}) \end{cases}$$

are symplectic integrators of order one. They are implicit for general Hamiltonian systems. For separable $H(q, p) = T(p) + U(q)$, however, both variants turn out to be explicit [15]. If the implicit midpoint rule is applied, then a second-order scheme is obtained:

$$(2.3) \qquad x^{j+1} = x^j + \delta t X_H \left(\frac{x^{j+1} + x^j}{2} \right).$$

In section 6, we will use the implicit midpoint rule for time integration. When X_H is a Hamiltonian vector field, (2.3) gives a symplectic scheme, and the sympletic structure is preserved at each step. When X_H is not a Hamiltonian vector field, (2.3) can still be used for time integration, but the energy and stability are not preserved in general.

Most of the usual numerical methods, such as the primitive Euler scheme and the classical Runge–Kutta scheme, are not symplectic integrators. A comprehensive review of symplectic integrators and their applications for Hamiltonian ODEs can be found in [15, 21]; the extension for Hamiltonian PDEs can be found in [5], where some structure-preserving discretization methods are discussed to transform Hamiltonian PDEs into Hamiltonian ODEs.

---

[1] The notations $[q, p]$ and $[q; p]$ are the same as the corresponding functions in MATLAB.

3. Symplectic projection. The symplectic projection takes advantage of empirical data to construct a reduced system while simultaneously preserving the underlying symplectic structure. In other words, if the original system is Hamiltonian, the reduced system remains Hamiltonian, but with significantly fewer dimensions.

3.1. Definitions of symplectic projection. Let $(\mathbb{V}, \Omega)$ and $(\mathbb{W}, \omega)$ be two symplectic vector spaces; $\dim(\mathbb{V}) = 2n$, $\dim(\mathbb{W}) = 2k$, and $k \leq n$.

DEFINITION 3.1. *A symplectic lift is a linear mapping $\sigma : \mathbb{W} \to \mathbb{V}$ that preserves the symplectic structure*

$$(3.1) \qquad \omega(z, w) = \Omega(\sigma(z), \sigma(w))$$

for all $z, w \in \mathbb{W}$.

Let $z \in \mathbb{W}$ and $x \in \mathbb{V}$. Using canonical coordinates, a symplectic lift $\sigma : z \mapsto x$ can be written as

$$x = Az,$$

where $A \in \mathbb{R}^{2n \times 2k}$, and satisfies

$$(3.2) \qquad A^T J_{2n} A = J_{2k}.$$

A matrix that satisfies (3.2) for some k and n with $k \leq n$ is called a *symplectic* matrix. The set of all $2n \times 2k$ symplectic matrices is the *symplectic Stiefel manifold*, denoted by $Sp(2k, \mathbb{R}^{2n})$. Moreover, since J_{2n} and J_{2k} are nonsingular, (3.2) itself requires that $k \leq n$ and $\text{rank}(A) = 2k$.

DEFINITION 3.2. *The* symplectic inverse *of a real matrix $A \in \mathbb{R}^{2n \times 2k}$, denoted as A^+, is defined by*

$$(3.3) \qquad A^+ = J_{2k}^T A^T J_{2n}.$$

Although A^+ is not equal to the Moore–Penrose pseudoinverse $(A^T A)^{-1} A^T$ in general, 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,*

$$(3.4) \qquad A = (A^+)^+,$$

$$(3.5) \qquad A = (((A^+)^T)^+)^T,$$

$$(3.6) \qquad A^+ J_{2n} = J_{2k} A^T.$$

LEMMA 3.4. *Suppose $A \in \mathbb{R}^{2n \times 2k}$ and A^+ is the symplectic inverse of A. Then the following are equivalent:*

(a) $A \in Sp(2k, \mathbb{R}^{2n})$.

(b) $(A^+)^T \in Sp(2k, \mathbb{R}^{2n})$.

(c) $A^+ A = I_{2k}$.

Proof. (a) $\Rightarrow$ (b). Replacing A^+ by (3.3) and using $A^T J_{2n} A = J_{2k}$ yield

$$A^+ J_{2n} (A^+)^T = (J_{2k}^T A^T J_{2n}) J_{2n} (J_{2n}^T A J_{2k}) = J_{2k}^T (A^T J_{2n} A) J_{2k} = J_{2k}.$$

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

(b) $\Rightarrow$ (c). Since $(A^+)^T \in Sp(2k, \mathbb{R}^{2n})$, we have $A^+ J_{2n} (A^+)^T = J_{2k}$. Substituting A by (3.5) and simplifying the expression yield

$$A^+ A = A^+ (((A^+)^T)^+)^T = A^+ (J_{2k}^T ((A^+)^T)^T J_{2n})^T = A^+ (J_{2k}^T A^+ J_{2n})^T$$
$$= A^+ J_{2n}^T (A^+)^T J_{2k} = -(A^+ J_{2n} (A^+)^T) J_{2k} = -J_{2k} J_{2k} = I_{2k}.$$

(c) $\Rightarrow$ (a). Replacing A^+ by (3.3) and plugging it into $A^+A = I_{2k}$, we obtain $J_{2k}^T A^T J_{2n} A = I_{2k}$. Left multiplying J_{2k} on both sides of this equation yields $A^T J_{2n} A = J_{2k}$. $\quad\square$

DEFINITION 3.5. *Suppose $\sigma : \mathbb{W} \to \mathbb{V}$ is a symplectic lift. Then the adjoint of σ is the linear mapping $\pi : \mathbb{V} \to \mathbb{W}$ satisfying*

$$(3.7) \qquad\qquad \omega(w, \pi(x)) = \Omega(\sigma(w), x)$$

for all $w \in \mathbb{W}$ and $x \in \mathbb{V}$. We say π is the symplectic projection *induced by σ.*

Since a symplectic projection is linear, $\pi : x \mapsto z$ has the form $z = Bx$, where $B \in \mathbb{R}^{2k \times 2n}$. Using canonical coordinates, (3.7) yields

$$w^T J_{2k} Bx = (Aw)^T J_{2n} x = w^T A^T J_{2n} x,$$

where A is the symplectic matrix that represents σ. Since w and x are arbitrary, we must have $J_{2k}B = A^T J_{2n}$. It follows that $B = A^+$, and the symplectic projection $\pi : x \mapsto z$ can be written as

$$(3.8) \qquad\qquad z = A^+ x.$$

Since $A^+A = I_{2k}$, $\pi \circ \sigma$ is the identity map on $\mathbb{W}$.

Remark 3.6. If we generalize $(\mathbb{W}, \Omega)$ and $(\mathbb{V}, \omega)$ to two symplectic manifolds and consider nonlinear transformations, the symplectic lift and symplectic projection respectively correspond to the symplectic embedding and symplectic submersion in symplectic geometry. Since this paper focuses on providing efficient numerical algorithms for practical applications, we consider only linear transformations between two vector spaces, although both the original and reduced systems can be nonlinear.

Now suppose $A^T J_{2n} A = J_{2k}$ and $x(t) \in \text{Range}(A)$ for all t. Then, $x(t) = Az(t)$. Using the chain rule, we obtain $\nabla_z H(Az) = A^T \nabla_x H(x)$. Taking the time derivative of (3.8) and using (2.2) and (3.6), the time evolution of z is given by

$$\dot{z} = A^+ \dot{x} = A^+ J_{2n} \nabla_x H(x) = J_{2k} A^T \nabla_x H(x) = J_{2k} \nabla_z H(Az),$$

where the last expression is a Hamiltonian vector field. Even if $x(t) \notin \text{Range}(A)$ for some t, the last expression is still well-defined. Thus, we can define the symplectic Galerkin projection.

DEFINITION 3.7. *The* symplectic Galerkin projection, *or* symplectic projection, *of a $2n$-dimensional Hamiltonian system $\dot{x} = J_{2n} \nabla_x H(x)$ with an initial condition $x(0) = x_0$ is given by a $2k$-dimensional ($k \leq n$) system*

$$(3.9) \qquad\qquad \dot{z} = J_{2k} \nabla_z \tilde{H}(z); \qquad z_0 = A^+ x_0,$$

where $\tilde{H} := H \circ A$ is the reduced Hamiltonian function, $A \in Sp(2k, \mathbb{R}^{2n})$ is a symplectic matrix, and $A^+ = J_{2k}^T A^T J_{2n}$ is the symplectic inverse of A.

Remark 3.8. Some Hamiltonian systems, such as the Burgers equation and the KdV equation, have nontrivial symplectic structures [20], which can be written in the form

$$(3.10) \qquad\qquad \dot{x} = J \nabla_x H(x),$$

where $J \in \mathbb{R}^{2n \times 2n}$ is a nonsingular skew-symmetric matrix. When $J \neq J_{2n}$, (3.10) does not denote a standard Hamiltonian system. Nevertheless, for any nonsingular skew-symmetric matrix J, there exists a congruent transformation such that

$QJQ^T = J_{2n}$ [11], where Q is a nonsingular matrix. Let $y = Qx$, then $\nabla_x H(x) = Q^T \nabla_y H(Q^{-1}y)$. It follows that

$$\dot{y} = Q\dot{x} = QJ\nabla_x H(x) = J_{2n}(Q^{-1})^T \nabla_x H(x) = J_{2n}\nabla_y H(Q^{-1}y).$$

The last equation indicates that a Hamiltonian equation with a nontrivial symplectic structure can be transformed to the canonical form and therefore can be simplified by the symplectic projection.

3.2. Linear Hamiltonian systems. A Hamiltonian system is *linear* if $H(x) = \frac{1}{2}x^T L x$, where L is a $2n \times 2n$ real symmetric matrix. Let $K := J_{2n}L$; the linear Hamiltonian system can be written as

$$\dot{x} = J_{2n}Lx = Kx. \tag{3.11}$$

A matrix of the form $K = J_{2n}L$, where L is symmetric, is called a *Hamiltonian matrix*. In addition, the set of all $2n \times 2n$ Hamiltonian matrices, denoted by $\mathfrak{sp}(\mathbb{R}^{2n})$, is a Lie algebra [22]. The fundamental matrix solution to (3.11) is given by

$$x(t) = e^{Kt}x_0. \tag{3.12}$$

Since $(\exp(Kt))^T J_{2n} \exp(Kt) = J_{2n}$, we have $\exp(tK) \in Sp(2n, \mathbb{R}^{2n})$, which means that the matrix exponential of a Hamiltonian matrix is symplectic. Conversely, the logarithm of a square symplectic matrix is Hamiltonian [23].

Plugging the $\tilde{H}(z) = H(Az) = \frac{1}{2}(Az)^T L(Az)$ into (3.9) yields a reduced system,

$$\dot{z} = J_{2k}\tilde{L}z = \tilde{K}z, \tag{3.13}$$

where $\tilde{L} := A^T L A$ and $\tilde{K} := J_{2k}\tilde{L}$. Since $\tilde{L}$ is symmetric, we have $\tilde{K} \in \mathfrak{sp}(\mathbb{R}^{2k})$, which implies that the reduced linear system (3.13) is also Hamiltonian.

Since the reduced system constructed by the symplectic projection is always Hamiltonian, energy and stability are preserved during the time evolution.

3.3. Energy preservation. Let $\Delta H(t) := H(x(t)) - \tilde{H}(z(t))$ denote the energy discrepancy between the state $x(t)$ and its approximation, $Az(t)$, derived from a reduced system. Since both the original and reduced systems are Hamiltonian, the system energy is conserved during time evolution. Moreover, $\tilde{H} = H \circ A$ by the definition. Thus, $\Delta H(t)$ is determined by the initial condition x_0 and the basis matrix A for all t, i.e.,

$$\Delta H(t) = H(x(0)) - \tilde{H}(z(0)) = H(x_0) - H(AA^+x_0). \tag{3.14}$$

If $x_0 \in \text{Range}(A)$, we have $AA^+x_0 = x_0$, which implies $\Delta H(t) = 0$ for all t; we say that the reduced system is *energy preserving*.

If $x_0 \notin \text{Range}(A)$, we can always extend A to a larger symplectic matrix A_{ext} such that the reduced system remains energy preserving. Specifically, suppose $A = [A_q, A_p]$ for $A_q, A_p \in \mathbb{R}^{2n \times k}$. Since $x_0 \notin \text{Range}(A)$, we must have $r_0 := x_0 - AA^+x_0 \neq 0$. Thus, the unit vector, $\hat{r}_0 := r_0/\|r_0\|$, is well-defined. One possible extension of A is given by

$$A_{\text{ext}} = [A_q, \hat{r}_0, A_p, J_{2n}^T \hat{r}_0]. \tag{3.15}$$

It is straightforward to verify that $A_{\text{ext}}^T J_{2n} A_{\text{ext}} = J_{2k+2}$ and $x_0 - A_{\text{ext}}A_{\text{ext}}^+x_0 = 0$. The last equation means that $x_0 \in \text{Range}(A_{\text{ext}})$, and therefore, $\Delta H(t) = H(x_0) - H(A_{\text{ext}}A_{\text{ext}}^+x_0) = 0$ for all t.

3.4. Stability preservation. In this subsection, we assume that the reduced system is constructed by the symplectic Galerkin projection. We also assume $x_0 \in$ Range(A), and the initial condition of the reduced system is given by $z_0 = A^+ x_0$. The following two theorems imply that energy preservation is a strong indicator for preserving stability.

THEOREM 3.9. *Consider the Hamiltonian system* (2.2) *with the initial condition* $x_0 \in \mathbb{R}^{2n}$. *If there exists a bounded neighborhood U of x_0 in $\mathbb{R}^{2n}$ such that $H(x_0) < H(x)$, or $H(x_0) > H(x)$, for all x on the boundary of U, then both the original system and the reduced system constructed by the symplectic projection are uniformly bounded for all t.*

Proof. We first assume that $H(x_0) < H(x)$ for all $x \in \partial U$, where ∂U denotes the boundary of U. Since U is bounded, so is ∂U. Because ∂U is also closed, ∂U is compact. Since $H : \mathbb{R}^{2n} \to \mathbb{R}$ is continuous, by the extreme value theorem, there exists a point x_1 of ∂U such that $H(x) \geq H(x_1)$ for all $x \in \partial U$. Since $H(x_0) < H(x)$ for all $x \in \partial U$, $H(x_0) < H(x_1)$. Because the system energy is conserved during time evolution, we have $H(x(t)) = H(x_0) < H(x_1)$ for all t. It follows that $x(t) \in U$ for all t, because if not, there is a time t_1 when $x(t_1) \in \partial U$, and $H(x(t_1)) \geq H(x_1)$, a contradiction.

Let $U_A = U \cap$ Range(A). Since U is a bounded open set in $\mathbb{V}$, U_A is also open in Range(A) and bounded. Let ∂U_A be the boundary of U_A in Range(A), $\partial U_A \subset \partial U \cap$ Range(A). Thus, $H(x_0) < H(x)$ for all $x \in U_A$. By the argument in the last paragraph, we have $Az(t) \in U_A$ for all t.

Finally, if $H(x_0) < H(x)$ is replaced by $H(x_0) > H(x)$, we can define $\hat{H}(x) = -H(x)$. Then, $\hat{H}(x_0) < \hat{H}(x)$ for all $x \in \partial U$. Thus, the conclusion still holds. □

An equilibrium point x_* of system (2.2) is *Lyapunov stable* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $\|x(t) - x_*\| < \epsilon$ for all $t > 0$ whenever $\|x_0 - x_*\| < \delta$. When (2.2) is linear and uniformly bounded, it is marginally stable in the sense of Lyapunov. In section 3.2, we show that when the original Hamiltonian system is linear, then the reduced system constructed by the symplectic projection is also linear. Thus, if the assumption of Theorem 3.9 holds, both the original and reduced systems are Lyapunov stable.

THEOREM 3.10. *If $x_* \in$ Range(A) is a strict local minimum or maximum of H, then x_* is a stable equilibrium for both the original Hamiltonian system and the reduced system constructed by the symplectic projection.*

Proof. We first assume that x_* is a strict local minimum of H. Then, there is an $\eta > 0$ such that $H(x_*) < H(x)$ for all x satisfying $0 < \|x - x_*\| \leq \eta$. Let $\kappa = \min(\epsilon, \eta)$ and $S = \{x \in \mathbb{R}^{2n} : \|x - x_*\| < \kappa\}$. Since ∂S is compact, by the extreme value theorem, there exists a point x_1 of ∂S such that $H(x) \geq H(x_1)$ for all $x \in \partial S$. Since $x_1 \in \partial S$, $H(x_*) < H(x_1)$. Because $H : \mathbb{R}^{2n} \to \mathbb{R}$ is continuous, there is a $\delta > 0$ such that $H(x) < H(x_1)$ for $\|x - x_*\| < \delta$. If $\|x_0 - x_*\| < \delta$, then $H(x_0) < H(x_1) \leq H(x)$ for all $x \in \partial S$. By Theorem 3.9, $x(t) \in S$ for all t. Thus, x_* is a stable equilibrium for (2.2).

Suppose U is a neighborhood of x_*, and x_* is the minimum of H in U. It immediately follows that x_* is also the minimum of H in U_A, where $U_A = U \cap$ Range(A). Thus, by the argument in the last paragraph, x_* is also the stable equilibrium of the reduced Hamiltonian system.

Finally, if x_* is a strict local maximum of H, x_* is a strict local minimum of $\hat{H}$, where $\hat{H}(x) = -H(x)$. Therefore, the conclusion still holds. □

The symplectic projection is analogous to the Galerkin projection, both of which construct reduced equations in some low-dimensional subspaces. However, the symplectic projection yields a reduced symplectic system by (3.9), while the Galerkin projection generally destroys the symplectic structure. Evolving the system (3.9) by a symplectic integrator preserves system energy and stability. By contrast, even if the POD subspace can provide an accurate representation of the empirical data, the reduced system constructed by the Galerkin projection may not be able to preserve these properties of the dynamics. In the next section, we shall discuss some PSD algorithms to construct a symplectic matrix A. This approach is an analogy to POD that constructs an orthonormal basis matrix.

4. Proper symplectic decomposition. Let $x(t_i) = [q(t_i); p(t_i)] \in \mathbb{R}^{2n}$ ($i = 1, \ldots, N$) denote N data points. Define a snapshot matrix

$$(4.1) \qquad M_x := [x(t_1), \ldots, x(t_N)].$$

The symplectic projection of M_x onto a low-dimensional subspace is given by $M_z = A^+ M_x$, where $A \in Sp(2k, \mathbb{R}^{2n})$, $M_z = [z(t_1), \ldots, z(t_N)] \in \mathbb{R}^{2k \times N}$, and $z(t_i) = A^+ x(t_i)$. The same projection of M_x in the original coordinates is given by $A M_z$, or $A A^+ M_x$.

The Frobenius norm $\| \cdot \|_F$ can be used to measure the error between M_x and its projection $\tilde{M}_x$. Suppose a symplectic matrix A minimizes the projection error in a least squares sense. Then, A is a solution of the following optimization problem:

$$(4.2) \qquad \begin{aligned} \text{minimize} \quad & \|M_x - A A^+ M_x\|_F \\ \text{subject to} \quad & A^T J_{2n} A = J_{2k}. \end{aligned}$$

Since the objective function has a fourth-order term in A after an expansion, (4.2) can be solved only iteratively. Because matrix A has $4nk$ elements, directly solving (4.2) is very expensive if $n \gg 1$. For this reason, we propose three efficient algorithms to construct an approximated optimal solution for the symplectic matrix A: these are the cotangent lift, complex SVD, and NLP.

4.1. Cotangent lift. In this section, an SVD-based algorithm is proposed to construct a symplectic matrix directly. 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. In particular, we define a set $\mathbb{M}_1(2n, 2k)$ by

$$(4.3) \qquad \mathbb{M}_1(2n, 2k) := Sp(2k, \mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \middle| \Phi \in \mathbb{R}^{n \times k} \right\}.$$

If $A_1 \in \mathbb{M}_1(2n, 2k)$, $A_1 = \mathrm{diag}(\Phi, \Phi)$ for some $\Phi \in \mathbb{R}^{n \times k}$. Then, $A_1^T J_{2n} A_1 = J_{2k}$ if and only if $\Phi^T \Phi = I_k$, which implies that Φ is an element of the Stiefel manifold $V_k(\mathbb{R}^n)$. It follows that

$$(4.4) \qquad \mathbb{M}_1(2n, 2k) = \left\{ \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \middle| \Phi \in V_k(\mathbb{R}^n) \right\}.$$

Let R and Q denote two vector spaces; $\dim(R) = k$, $\dim(Q) = n$, and $k \leq n$. Then, $TR \cong R \times R$ (resp., $TQ \cong Q \times Q$) gives the tangent bundle of R (resp., Q). Suppose $f : R \to Q$ and $\pi : Q \to R$ are linear mappings satisfying $\pi \circ f = \mathrm{id}_R$. Then, a *tangent lift* $f_* : TR \to TQ$ of f can be defined by $(r, w) \mapsto (f(r), f(w))$, and a

tangent lift $\pi_* : TQ \to TR$ of π can be defined by $(q, u) \mapsto (\pi(q), \pi(u))$. The f and the tangent lift π_* induce a cotangent lift π^* of π.

DEFINITION 4.1. *Let R^* (resp., Q^*) be the dual space of R (resp., Q), and let $\langle \cdot, \cdot \rangle_R := R^* \times R \to \mathbb{R}$ (resp., $\langle \cdot, \cdot \rangle_Q := Q^* \times Q \to \mathbb{R}$) be the duality paring. Let $\mathbb{W} = T^* R \cong R \times R^*$ (resp., $\mathbb{V} = T^* Q \cong Q \times Q^*$) denote the cotangent bundle of R (resp., Q). The* cotangent lift, $\pi^* : \mathbb{W} \to \mathbb{V}$, *of π is defined by a linear mapping*

$$(4.5) \qquad (r, s) \mapsto (f(r), \sigma(s)),$$

where $r \in R$, $s \in R^*$, and $\sigma : R^* \to Q^*$ is the adjoint of π satisfying

$$(4.6) \qquad \langle \sigma(s), u \rangle_Q = \langle s, \pi(u) \rangle_R$$

for all $u \in Q$.

Given bases in R and Q, the linear mappings f and π can be represented by matrices Φ and Ψ^T for $\Phi, \Psi \in \mathbb{R}^{n \times k}$. The constraint $\pi \circ f = \mathrm{id}_R$ requires that $\Psi^T \Phi = I_k$. By choosing dual bases in R^* and Q^*, (4.6) yields $\sigma(s) = \Psi s$. Thus, the cotangent lift π^* can be written as

$$x = \pi^*(z) = Bz$$

for $z \in \mathbb{W}$ and $x \in \mathbb{V}$, where $B := \mathrm{diag}(\Phi, \Psi)$. Since $B^+ B = I_{2k}$, Lemma 3.4 yields that $B \in Sp(2k, \mathbb{R}^{2n})$.

Especially when $\Phi = \Psi$, B degenerates to A_1, and the constraint $\Psi^T \Phi = I_k$ becomes $\Phi^T \Phi = I_k$. In this scenario, the range of Φ should approximately fit for both $q(t)$ and $p(t)$. As Algorithm 1 indicates, Φ can be computed by the SVD of an extended snapshot matrix $M_1 \in \mathbb{R}^{n \times 2N}$, which is defined by

$$(4.7) \qquad M_1 := [q(t_1), \ldots, q(t_N), \gamma p(t_1), \ldots, \gamma p(t_N)],$$

where γ represents a weighting coefficient that balances the SVD truncation error of $q(t_i)$ with $p(t_i)$. Let $\hat{q}(t)$ and $\hat{p}(t)$ denote approximating solutions obtained from a reduced system. If the goal is to minimize $\|[\hat{q}(t); \hat{p}(t)] - [q(t); p(t)]\|_2$, we can choose $\gamma = 1$ so that $q(t)$ and $p(t)$ have the same status in the SVD truncation. If the goal is to minimize $\|\hat{q}(t) - q(t)\|_2$, we can simply choose $\gamma = 0$.[2]

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 block-diagonal form.
  1: Construct an extended snapshot matrix M_1 as (4.7).
  2: Compute the SVD of M_1 to obtain a POD basis matrix Φ.
  3: Construct the symplectic matrix $A_1 = \mathrm{diag}(\Phi, \Phi)$.

THEOREM 4.2. *Suppose $M_x \in \mathbb{R}^{2n \times N}$ is the snapshot matrix defined by (4.1). If we select $\gamma = 1$ in (4.7), the symplectic matrix A_1 constructed by Algorithm 1 is the optimal solution in $\mathbb{M}_1(2n, 2k)$ that minimizes the error in the projection of M_x onto the column space.*

---

[2]However, if the Hamiltonian has the form $H(q, p) = \frac{1}{2} p^T p - V(q)$, where $V : \mathbb{R}^n \to \mathbb{R}$ denotes the potential function, one may consider $q(t_i) \pm \delta t \dot{p}(t_i)$ to also be a snapshot of the trajectory of $q(t)$. Thus, if the goal is to minimize $\|\hat{q}(t) - q(t)\|_2$, we can also choose $\gamma = \delta t$ for this special case.

Proof. Similar to (4.2), we can express the optimization problem as

$$
\begin{aligned}
\text{minimize} \quad & \|M_x - A_1 A_1^+ M_x\|_F \\
\text{subject to} \quad & A_1 \in \mathbb{M}_1(2n, 2k).
\end{aligned}
\tag{4.8}
$$

Let $M_q := [q(t_1), \ldots, q(t_N)]$ and $M_p := [p(t_1), \ldots, p(t_N)]$. By definition, $M_x = [M_q; M_p]$. Since $A_1 = \text{diag}(\Phi, \Phi)$ and $\Phi^T \Phi = I_k$, we have $A^+ = \text{diag}(\Phi^T, \Phi^T)$. Moreover, $\gamma = 1$ implies that $M_1 = [M_q, M_p]$. Then, the objective function becomes

$$
\begin{aligned}
\|M_x - A_1 A_1^+ M_x\|_F &= \left\| \begin{bmatrix} M_q \\ M_p \end{bmatrix} - \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \begin{bmatrix} \Phi^T & 0 \\ 0 & \Phi^T \end{bmatrix} \begin{bmatrix} M_q \\ M_p \end{bmatrix} \right\|_F \\
&= \left\| \begin{bmatrix} (I_n - \Phi\Phi^T)M_q \\ (I_n - \Phi\Phi^T)M_p \end{bmatrix} \right\|_F \\
&= \left\| [(I_n - \Phi\Phi^T)M_q, \ (I_n - \Phi\Phi^T)M_p] \right\|_F \\
&= \left\| (I_n - \Phi\Phi^T)[M_q, \ M_p] \right\|_F \\
&= \left\| M_1 - \Phi\Phi^T M_1 \right\|_F.
\end{aligned}
$$

Thus, Φ can be directly solved by the truncated SVD of M_1,

$$
M_1 \approx \Phi \Sigma \Psi^T,
\tag{4.9}
$$

where the matrix Σ is a $k \times k$ diagonal matrix with nonnegative real numbers on the diagonal; Φ and Ψ are real matrices and satisfy $\Phi^T \Phi = \Psi^T \Psi = I_k$. Thus, the symplectic matrix A_1 constructed by Algorithm 1 is the optimal solution for the optimization problem (4.8). ☐

It should be mentioned that in [18], 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 $\Phi \in \mathbb{R}^{n \times k}$ can be constructed by the SVD of a snapshot matrix $[q(t_1), \ldots, q(t_N)]$ for $q(t) \in Q$. Then, 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 TR, where $r(t) \in R$. Thus, a reduced system for $(r, \dot{r})$ can be given by the Euler–Lagrange equation of $\tilde{L}(r, \dot{r})$.

By the Legendre transformation, the reduced Lagrangian system can be transformed into a reduced Hamiltonian system. Meanwhile, the proposed cotangent lift method can yield another reduced Hamiltonian system. However, the two reduced systems are not equal in general, in two aspects.

First, the two reduced systems reside on different subspaces. In either case, the subspace can be presented as the column space of $A_1 = \text{diag}(\Phi, \Phi)$, where Φ is a POD basis matrix for the generalized coordinates. In [18], the tangent lift constructs Φ from a snapshot ensemble of $q(t)$. In this paper, the proposed cotangent lift constructs Φ from a snapshot ensemble of $q(t)$ and $p(t)$, as (4.7) indicates.

Second, the two reduced systems give different trajectories on $\mathbb{R}^{2k}$. Consider a Lagrangian of the form $L(q, \dot{q}) = \frac{1}{2} \dot{q}^T M \dot{q} - V(q)$, where the mass matrix $M \in \mathbb{R}^{n \times n}$ is nonsingular and the potential function $V : \mathbb{R}^n \to \mathbb{R}$ is smooth. Then, the reduced Lagrangian is given by $\tilde{L}(r, \dot{r}) = \frac{1}{2} \dot{r}^T \tilde{M} \dot{r} - V(\Phi r)$, where $\tilde{M} = \Phi^T M \Phi \in \mathbb{R}^{k \times k}$. Suppose $\tilde{M}$ is also nonsingular; the Legendre transform produces the reduced momentum by $s = \tilde{M} \dot{r}$ and the corresponding Hamiltonian by

$$
\tilde{H}_1(r, s) = s^T \dot{r} - \tilde{L}(r, \dot{r}) = \frac{1}{2} s^T \tilde{M}^{-1} s + V(\Phi r).
$$

On the other hand, with $p = M\dot{q}$, the original Hamiltonian in $\mathbb{R}^{2n}$ is $H(q, p) = \frac{1}{2}p^T M^{-1}p + V(q)$. Then, the reduced Hamiltonian constructed by the cotangent lift method on $\mathbb{R}^{2k}$ is given by

$$\tilde{H}(r, s) = H(\Phi r, \Phi s) = \frac{1}{2}s^T(\Phi^T M^{-1}\Phi)s + V(\Phi r).$$

Unless $M = I_n$, $\Phi^T M^{-1}\Phi = \tilde{M}^{-1}$ does not hold in general; thus $\tilde{H}(r, s)$ and $\tilde{H}_1(r, s)$ can determine two different trajectories on $\mathbb{R}^{2k}$.

4.2. Complex singular value decomposition. This section proposes an SVD-based algorithm to construct a symplectic basis matrix, such that the off-diagonal blocks are nonzero submatrices. If we use $q(t) + \iota p(t)$ to describe the solution trajectory in the phase space ($\iota = \sqrt{-1}$), we can construct a complex snapshot matrix $M_2 \in \mathbb{C}^{n \times N}$ by

$$(4.10) \qquad M_2 := [q(t_1) + \iota p(t_1), \ldots, q(t_N), +\iota p(t_N)].$$

By definition, we have $M_2 = M_q + \iota M_p$. Suppose a unitary matrix $U \in \mathbb{C}^{n \times k}$ minimizes the error in the projection of M_2 onto the column space. Then, U can be obtained from the following optimization problem:

$$(4.11) \qquad \begin{aligned} &\text{minimize} \quad \left\| M_2 - UU^H M_2 \right\|_F \\ &\text{subject to} \quad U^H U = I_k, \end{aligned}$$

where U^H is the conjugate transpose of U. The optimal value of U can also be obtained by the truncated SVD of M_2,

$$(4.12) \qquad M_2 \approx U\Sigma V^H,$$

where the matrix Σ 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 a complex Stiefel manifold. Then, its element $U \in V_k(\mathbb{C}^n)$ has the form $U = \Phi + \iota\Psi$, where $\Phi, \Psi \in \mathbb{R}^{n \times k}$. We define a mapping $\mathscr{A} : V_k(\mathbb{C}^n) \to \mathbb{R}^{2n \times 2k}$ by the formula

$$(4.13) \qquad \mathscr{A}(U) = \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix}.$$

LEMMA 4.3. *The mapping $\mathscr{A}$ is injective. The image of $\mathscr{A}$ is equal to $\mathbb{M}_2(2n, 2k)$, where*

$$(4.14) \qquad \mathbb{M}_2(2n, 2k) := Sp(2k, \mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix} \middle| \Phi, \Psi \in \mathbb{R}^{n \times k} \right\}.$$

Proof. It follows from $\mathscr{A}$'s definition that it is injective. If $\Phi + \iota\Psi \in V_k(\mathbb{C}^n)$, then $(\Phi + \iota\Psi)^H(\Phi + \iota\Psi) = I_k$, which is equivalent to

$$(4.15) \qquad \Phi^T \Phi + \Psi^T \Psi = I_k, \quad \Phi^T \Psi = \Psi^T \Phi.$$

Let $A_2 = \mathscr{A}(\Phi + \iota\Psi)$. Using (4.15), it is easy to verify that $A_2^T J_{2n} A_2 = J_{2k}$. Thus, $A_2 \in \mathbb{M}_2(2n, 2k)$, i.e., $\mathscr{A}(V_k(\mathbb{C}^n)) \subset \mathbb{M}_2(2n, 2k)$.

Conversely, if $A_2 \in \mathbb{M}_2(2n, 2k)$, then $A_2^T J_{2n} A_2 = J_{2k}$. Plugging $A_2 = [\Phi, -\Psi; \Psi, \Phi]$ into $A_2^T J_{2n} A_2 = J_{2k}$ gives (4.15). It follows that $(\Phi + \iota\Psi)^H (\Phi + \iota\Psi) = I_k$. As a result, $\Phi + \iota\Psi \in V_k(\mathbb{C}^n)$, and $\mathscr{A}^{-1}(\mathbb{M}_2(2n, 2k)) \subset V_k(\mathbb{C}^n)$. □

Lemma 4.3 implies that the function $\mathscr{A}' : V_k(\mathbb{C}^n) \to \mathbb{M}_2(2n, 2k)$ obtained by restricting the range of $\mathscr{A}$ is bijective. A topology on $V_k(\mathbb{C}^n)$ (resp., $\mathbb{M}_2(2n, 2k)$) can be set as the subspace topology of $\mathbb{C}^{n \times k}$ (resp., $\mathbb{R}^{2n \times 2k}$) that is induced by a matrix norm. It is easy to show that both $\mathscr{A}'$ and $(\mathscr{A}')^{-1}$ are continuous mappings in terms of the subspace topology. Thus, $\mathscr{A}'$ is a homeomorphism, and $\mathbb{M}_2(2n, 2k)$ is a submanifold of $Sp(2k, \mathbb{R}^{2n})$. Consequently, a symplectic matrix A_2 can be constructed through the mapping $\mathscr{A}$. Algorithm 2 outlines the procedure.

ALGORITHM 2. COMPLEX SINGULAR VALUE DECOMPOSITION.

Require: An empirical data ensemble $\{q(t_i), p(t_i)\}_{i=1}^N$.
Ensure: A symplectic matrix A_2 in block form.
  1: Construct a complex snapshot matrix M_2 as (4.10).
  2: Compute the SVD of M_2 to obtain a basis matrix $\Phi + \iota\Psi$.
  3: Construct the symplectic matrix $A_2 = [\Phi, -\Psi; \Psi, \Phi]$.

Matrices in $\mathbb{M}_2(2n, 2k)$ are not only symplectic but also orthonormal. To see this, we can substitute $A_2 = [\Phi, \Psi; -\Psi, \Phi]$ into (3.3) and obtain $A_2^+ = J_{2k}^T A_2^T J_{2n} = A_2^T$. It follows that $A_2^T A_2 = A_2^+ A_2 = I_{2k}$, i.e., $A_2 \in V_{2k}(\mathbb{R}^{2n})$. Conversely, for any $A_2 = [\Phi, \Psi; -\Psi, \Phi]$ that belongs to $V_{2k}(\mathbb{R}^{2n})$, (4.15) holds, which means that $A_2 \in Sp(2k, \mathbb{R}^{2n})$. Therefore, $\mathbb{M}_2(2n, 2k)$ can also be defined by

$$(4.16) \qquad \mathbb{M}_2(2n, 2k) = V_{2k}(\mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix} \middle| \Phi, \Psi \in \mathbb{R}^{n \times k} \right\}.$$

The following lemma gives the other equivalent definition of $\mathbb{M}_2(2n, 2k)$.

LEMMA 4.4. *Suppose $\mathbb{M}_2(2n, 2k)$ is the matrix manifold of $\mathbb{R}^{2n \times 2k}$ defined by* (4.14). *Then,*

$$(4.17) \qquad \mathbb{M}_2(2n, 2k) = Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n}).$$

Proof. By (4.14) and (4.16), $\mathbb{M}_2(2n, 2k) \subset Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n})$. Conversely, let $A_2 = [A_q, A_p] \in Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n})$, where $A_q = [\xi_1, \ldots, \xi_k] \in \mathbb{R}^{2n \times k}$ and $A_p = [\zeta_1, \ldots, \zeta_k] \in \mathbb{R}^{2n \times k}$. Orthonormality of A_2 requires that $\|\xi_i\| = \|\zeta_i\| = 1$ for any $i \in \{1, \ldots, k\}$. It follows that $\|J_{2n}^T \xi_i\| = 1$. Thus, $\|J_{2n}^T \xi_i\| = \|\zeta_i\|$. On the other hand, symplecticity of A_2 requires that $\Omega(\xi_i, \zeta_i) = 1$, which implies that $\langle J_{2n}^T \xi_i, \zeta_i \rangle = 1$. The Cauchy–Schwarz inequality states that $\langle J_{2n}^T \xi_i, \zeta_i \rangle \leq \|J_{2n}^T \xi_i\| \cdot \|\zeta_i\|$, and two sides are equal if and only if $J_{2n}^T \xi_i$ and ζ_i are parallel. The combination of orthonormality and symplecticity gives $J_{2n}^T \xi_i = \zeta_i$. It follows that A_2 must have the block form $[A_q, J_{2n}^T A_q]$, or $[\Phi, -\Psi; \Psi, \Phi]$ if A_q is written as $[\Phi; \Psi]$ for $\Phi, \Psi \in \mathbb{R}^{n \times k}$. Therefore, $A_2 \in \mathbb{M}_2(2n, 2k)$. Since A_2 is arbitrary, $Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n}) \subset \mathbb{M}_2(2n, 2k)$. □

Since $\mathscr{A}'$ is a homeomorphism, (4.16) and (4.17) yield

$$V_k(\mathbb{C}^n) \cong \mathbb{M}_2(2n, 2k) = V_{2k}(\mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & -\Psi \\ \Psi & \Phi \end{bmatrix} \middle| \Phi, \Psi \in \mathbb{R}^{n \times k} \right\}$$

$$(4.18) \qquad\qquad\qquad = Sp(2k, \mathbb{R}^{2n}) \cap V_{2k}(\mathbb{R}^{2n}).$$

Specifically, when $\Psi = 0$, we obtain

$$(4.19) \qquad V_k(\mathbb{R}^n) \cong \mathbb{M}_1(2n, 2k) = V_{2k}(\mathbb{R}^{2n}) \cap \left\{ \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \middle| \Phi \in \mathbb{R}^{n \times k} \right\}.$$

Moreover, $\mathscr{A}$ also preserves algebraic structures, as one can easily verify the following lemma.

LEMMA 4.5. *Let $C \in \mathbb{C}^{n_1 \times n_2}$ and $D \in \mathbb{C}^{n_2 \times n_3}$. Then, we have $\mathscr{A}(C)\mathscr{A}(D) = \mathscr{A}(CD)$ and $\mathscr{A}(C^H) = (\mathscr{A}(C))^T$.*

THEOREM 4.6. *Suppose $M_x \in \mathbb{R}^{2n \times N}$ is the snapshot matrix defined by (4.1). The symplectic matrix A_2 constructed by Algorithm 2 is an optimal solution in $\mathbb{M}_2(2n, 2k)$ that minimizes the error in the projection of $[M_x, J_{2n}^T M_x]$ onto the column space.*

Proof. By Lemma 4.5, the truncated SVD of M_2 given by (4.12) yields

$$(4.20) \qquad \mathscr{A}(M_2) \approx \mathscr{A}(U)\mathscr{A}(\Sigma)\mathscr{A}(V^H) = \mathscr{A}(U)\mathscr{A}(\Sigma)(\mathscr{A}(V))^T.$$

Since $U^H U = I_k$, by Lemma 4.5, we have

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

Similarly, $(\mathscr{A}(V))^T \mathscr{A}(V) = I_{2k}$ holds due to $V^H V = I_k$. Moreover, $\mathscr{A}(\Sigma)$ is a real diagonal matrix that contains the largest $2k$ singular values of $\mathscr{A}(M_2)$. Thus, the right-hand side of (4.20) provides a truncated SVD for $\mathscr{A}(M_2)$.

In Algorithm 2, the symplectic matrix is constructed by $A_2 = \mathscr{A}(U)$. Meanwhile, using the definition of M_2 and M_x, we have

$$\mathscr{A}(M_2) = \begin{bmatrix} M_q & -M_p \\ M_p & M_q \end{bmatrix} = \begin{bmatrix} M_x, & J_{2n}^T M_x \end{bmatrix}.$$

Thus, A_2 is a POD basis matrix of $[M_x, J_{2n}^T M_x]$. Consequently, A_2 is an optimal matrix in $\mathbb{R}^{2n \times 2k}$ (more than just in $\mathbb{M}_2(2n, 2k)$) that minimizes the projection error of $[M_x, J_{2n}^T M_x]$. $\square$

Theorem 4.6 implies that the complex SVD is designed to fit $[M_x, J_{2n}^T M_x]$, rather than M_x itself. As a result, Algorithm 2 can only construct a near optimal matrix in $\mathbb{M}_2$.

4.3. Nonlinear programming. Although it is often too expensive to solve the optimization problem (4.2) directly, one can search a near optimal solution over a subset of $Sp(2k, \mathbb{R}^{2n})$. The proposed NLP algorithm here is analogous to the optimization algorithm in [12], where an optimal POD basis matrix is constructed from a linear transformation of a snapshot matrix.

In particular, if we have a prespecified basis matrix $A_1 \in \mathbb{M}_1(2n, 2r)$ or $A_2 \in \mathbb{M}_2(2n, 2r)$, with $k \le r \le n$, we may assume that the near optimal solution $A_3 \in Sp(2k, \mathbb{R}^{2n})$ is a linear transformation of A_1 or A_2. Now suppose the cotangent lift yields a symplectic matrix A_1 in $\mathbb{M}_1(2n, 2r)$. If $\mathrm{Range}(A_3) \subset \mathrm{Range}(A_1)$, we have

$$(4.21) \qquad A_3 = A_1 \cdot C,$$

where $C \in \mathbb{R}^{2r \times 2k}$ is the coefficient matrix of A_3 with respect to the basis vectors of A_1. Plugging (4.21) into $A_3^T J_{2n} A_3 = J_{2k}$ and using $A_1^T J_{2n} A_1 = J_{2r}$ give

$$(4.22) \qquad C^T J_{2r} C = J_{2k},$$

which implies $C \in Sp(2k, \mathbb{R}^{2r})$. Since $(A_1 C)^+ = C^+ A_1^+$, the original optimization problem (4.2) reduces to

$$(4.23) \qquad \begin{aligned} &\text{minimize} \quad \| M_x - A_1 C C^+ A_1^+ M_x \|_F \\ &\text{subject to} \quad C^T J_{2r} C = J_{2k}. \end{aligned}$$

Let $A_1 = \mathrm{diag}(\Phi, \Phi)$, where $\Phi \in \mathbb{R}^{n \times r}$. The initial value for (4.2) can be $A_3 = \mathrm{diag}(\Phi', \Phi')$, where Φ' denotes the first k columns of Φ. Correspondingly, the initial value for (4.23) is given by $C = \mathrm{diag}(I_{r \times k}, I_{r \times k})$, where $I_{r \times k}$ denotes the first k columns of the identity matrix I_r. When $r \ll n$, the size of C is much smaller than the size of A_3, and thus the computational cost of (4.23) is significantly lower than the original optimization problem (4.2).

ALGORITHM 3. NONLINEAR PROGRAMMING.

Require: An empirical data ensemble $\{q(t_i), p(t_i)\}_{i=1}^N$.
Ensure: A symplectic matrix $A_3 \in Sp(2k, \mathbb{R}^{2n})$.
  1: Construct a symplectic matrix $A_1 \in \mathbb{M}_1(2n, 2r)$ with $r > k$ by the cotangent lift.
  2: Solve (4.23) and obtain a coefficient matrix $C \in Sp(2k, \mathbb{R}^{2r})$.
  3: Construct the symplectic matrix $A_3 = A_1 \cdot C$.

So far, three different PSD algorithms have been proposed to construct a symplectic basis matrix. Corresponding to three manifolds with the inclusion maps,

$$V_k(\mathbb{R}^n) \hookrightarrow V_k(\mathbb{C}^n) \xrightarrow{\mathscr{A}} Sp(2k, \mathbb{R}^{2n}),$$

we propose the cotangent lift, complex SVD, and NLP. The cotangent lift and complex SVD algorithms are faster and more easily implemented in offline computation; their computational costs only involve the SVD. However, both algorithms search optimal basis matrices in submanifolds of $Sp(2k, \mathbb{R}^{2n})$, rather than in $Sp(2k, \mathbb{R}^{2n})$ itself. Therefore, they sacrifice certain accuracy to fit the empirical data in order to reduce costs. By contrast, the NLP is more expensive in offline computation, since it requires solving an optimization problem in $Sp(2k, \mathbb{R}^{2r})$ based on a prespecified basis matrix constructed by another algorithm. However, the NLP can result in a symplectic matrix to fit the empirical data with less projection error.

For the cotangent lift, we have a parameter γ in (4.7) to balance the truncation error of $q(t)$ and $p(t)$. At first glance, the other two algorithms do not have a similar weighting option. However, we can always construct a linear transformation from $x = [q; p]$ to $x_\gamma = [q; p_\gamma]$ by $p_\gamma = \gamma p$ and then solve the rescaled Hamiltonian equation based on $\tilde{H}(q, p_\gamma) := H(q, p_\gamma/\gamma)$. The fully resolved rescaled Hamiltonian system is equivalent to the original one; depending on the subspace on which the reduced system lives, however, the reduced models for the original and rescaled systems are not equivalent in general. A weighted data ensemble for the rescaled system can be defined as

$$(4.24) \qquad M_{x_\gamma} := [x_\gamma(t_1), \ldots, x_\gamma(t_N)].$$

Then, a symplectic subspace can be constructed to fit M_{x_γ} by any of the aforementioned PSD algorithms. Thus, the complex SVD and NLP can also flexibly balance the truncation error of $p(t)$ and $q(t)$ by choosing a suitable value of γ.

5. Symplectic model reduction of nonlinear Hamiltonian systems. As an approximation of the symplectic Galerkin projection, SDEIM is developed in this section. The motivation of SDEIM is to reduce the computational complexity of a nonlinear Hamiltonian system while simultaneously preserving the symplectic structure. Before introducing SDEIM, we will give a review of the classical DEIM.

5.1. Discrete empirical interpolation method. Let $x \in \mathbb{R}^n$ denote the state variable in the original space and let $f : \mathbb{R}^n \to \mathbb{R}^n$ denote the discretized vector field. The full-order dynamical system can be described by an initial value problem

$$(5.1) \qquad \dot{x} = f(x) = Lx + f_N(x); \quad x(0) = x_0,$$

where the original vector field $f(x)$ is split into a linear part Lx with $L \in \mathbb{R}^{n \times n}$ and a nonlinear part $f_N(x)$ with $f_N : \mathbb{R}^n \to \mathbb{R}^n$.

Let $\Phi \in \mathbb{R}^{n \times k}$ denote a POD basis matrix. Then, the Galerkin projection can be used to obtain a reduced system on the column space of Φ,

$$(5.2) \qquad \dot{z} = \Phi^T f(\Phi z) = \tilde{L}z + \Phi^T f_N(\Phi z); \quad z_0 = \Phi^T x_0,$$

where $z(t) \in \mathbb{R}^k$ is the reduced state, and $\tilde{L} = \Phi^T L \Phi \in \mathbb{R}^{k \times k}$ is the reduced linear operator.

According to some previous studies [9, 27], the POD-Galerkin approach can achieve computational savings only when the analytical formula of the nonlinear vector term $\Phi^T f_N(\Phi z)$ can be simplified, especially if $f_N(x)$ is a low-order polynomial in x. Otherwise, one usually needs to compute the state variable $x := \Phi z$ in the original coordinate system, evaluate the nonlinear vector field $f_N(x)$, and then project $f_N(x)$ back to the column space of Φ. In this scenario, solving the POD reduced system can be more expensive than solving the original full-order system.

DEIM focuses on approximating $f_N(x)$ so that a certain coefficient matrix can be precomputed and, as a result, the complexity in evaluating $f_N(x)$ becomes proportional to the small number of selected spatial indices [9]. Let $\beta = [\beta_1; \ldots; \beta_m] \in \mathbb{R}^m$ be an index vector, and let $\beta_i \in \{1, \ldots, n\}$ be an index. Define an $n \times m$ matrix

$$(5.3) \qquad P := [e_{\beta_1}, \ldots, e_{\beta_m}],$$

where e_{β_i} is the β_ith column of the identity matrix I_{2n}. Then, left multiplication of $f_N(x)$ with P^T projects $f_N(x)$ onto m elements corresponding to the index vector β. Now suppose $f_N(x)$ resides approximately on the range of an $n \times m$ matrix Ψ; then there exists a corresponding coefficient vector $\tau \in \mathbb{R}^m$ such that $f_N(x) \approx \Psi\tau$. The coefficient vector τ can be determined by matching the $f_N(x)$ at selected m spatial indices, i.e., $P^T f_N(x) = P^T \Psi\tau$. Suppose $P^T f_N(x)$ is nonsingular. Then, we have $\tau = (P^T \Psi)^{-1} P^T f_N(x)$. Thus, the approximation $\hat{f}_N(x)$ of the nonlinear vector term $f_N(x)$ becomes

$$(5.4) \qquad \hat{f}_N(x) = \Psi\tau = \Psi(P^T \Psi)^{-1} P^T f_N(x),$$

and the reduced system (5.2) can be approximated as

$$(5.5) \qquad \dot{z} = \tilde{L}z + Wg(z),$$

where $W = \Phi^T \Psi(P^T \Psi)^{-1}$ and $g(z) = P^T f_N(\Phi z)$. Notice that W is calculated only once at the offline stage. At the online stage, $g(z)$ is evaluated on m spatial indices of $f_N(\Phi z)$. Therefore, the complexity of the DEIM-reduced system (5.5) can be independent of the dimension of the original system.

In order to construct (5.5), the SVD can be applied to construct a POD basis matrix Φ based on an empirical data ensemble $[x(t_1), \ldots, x(t_N)]$ and the collateral POD basis matrix Ψ based on another data ensemble $[f_N(x(t_1)), \ldots, f_N(x(t_N))]$ for the nonlinear vector term. Moreover, a greedy algorithm can be applied to construct

---

ALGORITHM 4. GREEDY ALGORITHM TO CONSTRUCT AN INDEX VECTOR β.

Require: A basis matrix $\Psi = [\psi_1, \ldots, \psi_m] \in \mathbb{R}^{n \times m}$.
Ensure: An index vector $\beta = [\beta_1; \ldots; \beta_m] \in \mathbb{R}^m$.
  1: Select the first interpolation index $[\rho, \beta_1] = \max\{|\psi_1|\}$.
  2: Initialize $U = [\psi_1]$, $\beta = \beta_1$.
for $i = 2$ to m **do**
    3: Solve the coefficient vector τ to match ψ_i, $U(\beta, :)\tau = \psi_i(\beta)$.
    4: Calculate the residual $r = \psi_i - U\tau$.
    5: Select the interpolation index corresponding to the largest magnitude of the
    residual r, $[\rho, \beta_i] = \max\{|r|\}$.
    6: Update $U = [U, \ \psi_i]$, $\beta = [\beta; \beta_i]$.
end for

---

the index vector β [9], as listed in Algorithm 4.[3] Initially, we select the first inter-
polation index $\beta_1 \in \{1, \ldots, n\}$ corresponding to the first basis function ψ_1 with the
largest magnitude. The remaining interpolation indices, β_i for $i = 2, \ldots, m$, respec-
tively correspond to the largest magnitude of the residual r, where r is the residual
between the input basis ψ_i and its projection onto the column space of U. In step 5,
$[\rho, \beta_i] = \max\{|r|\}$ means $\rho = |r(\beta_i)| = \max_{j=1,\ldots,n}|r(j)|$. In step 6, we add a column
vector ψ_i (and an element β_i) to a matrix U (and a vector β). It has been proven
that $\rho \neq 0$ implies that $P^T A$ is nonsingular [9]. Thus, the DEIM approximation of
the nonlinear vector term $\hat{f}_N(x)$ in (5.4) is well-defined.

5.2. Symplectic discrete empirical interpolation method. In this section,
we present a method that combines the idea of symplectic Galerkin projection with
the use of DEIM. Similar to (5.1), the original Hamiltonian can also be split into a
linear part and a nonlinear part, i.e., $H(x) = H_1(x) + H_2(x)$, such that $\nabla_x H_1(x) = Lx$
for a real symmetric matrix L, and $\nabla_x H_2(x) = f_N(x)$ for a nonlinear function f_N.
Thus, the original Hamiltonian system is given by

$$(5.6) \qquad \dot{x} = J_{2n}\nabla_x H(x) = Kx + J_{2n}f_N(x),$$

where $K = J_{2n}L \in \mathfrak{sp}(\mathbb{R}^{2n})$. Analogous to (5.2), the symplectic Galerkin projection
yields the following reduced Hamiltonian system:

$$(5.7) \qquad \dot{z} = A^+ J_{2n}\nabla_x H(Az) = \tilde{K}z + J_{2k}A^T f_N(Az),$$

where $A \in Sp(2k, \mathbb{R}^{2n})$, $\tilde{L} = A^T LA$, and $\tilde{K} = A^+ KA = J_{2k}\tilde{L}$. Thus, unless
$A^T f_N(Az)$ can be analytically simplified, the computational complexity of (5.7) still
depends on $2n$. In order to save the computational cost, one can use the DEIM ap-
proximation (5.4) to approximate the nonlinear vector term $f_N(x)$. Let $\Psi \in \mathbb{R}^{2n \times m}$
denote the collateral POD basis for $f_N(x)$, and let $P \in \mathbb{R}^{2n \times m}$ denote the projection
matrix with the form (5.3). Then, (5.7) can be approximated as

$$(5.8) \qquad \dot{z} = \tilde{K}z + J_{2k}Wg(z),$$

where $W = A^T \Psi(P^T \Psi)^{-1}$ and $g(z) = P^T f_N(Az)$.

---

[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).

THEOREM 5.1. *Let $z(t)$ be the solution of* (5.8) *with* $z(0) = z_0$. *Let* $E(t) = \tilde{H}(z(t))$ *be the corresponding system energy. Then, the time derivative of $E(t)$ equals*

$$(5.9) \qquad \dot{E}(t) = (\nabla_z \tilde{H}(z))^T J_{2k} A^T (\hat{f}_N(Az) - f_N(Az)),$$

where $\hat{f}_N$ is given by (5.4). *Moreover, an upper bound to $\dot{E}(t)$ is given by*

$$(5.10) \qquad \|\dot{E}(t)\| \le C \|\nabla_z \tilde{H}(z)\| \cdot \|(I - \Psi\Psi^T) f_N(Az)\|,$$

where $C = \|(P^T \Psi)^{-1}\|$ is a constant.

Proof. Taking the time derivative of $E(t)$ and using the chain rule gives $\dot{E}(t) = (\nabla_z \tilde{H}(z))^T \dot{z}$. Replacing $\dot{z}$ by (5.8) yields

$$\dot{E}(t) = (\nabla_z \tilde{H}(z))^T J_{2k}(\tilde{L}z + Wg(z)).$$

Because J_{2k} is skew-symmetric, we have $(\nabla_z \tilde{H}(z))^T J_{2k} \nabla_z \tilde{H}(z) = 0$. It follows that

$$\dot{E}(t) = (\nabla_z \tilde{H}(z))^T J_{2k}(\tilde{L}z + Wg(z) - \nabla_z \tilde{H}(z)).$$

Thus, (5.9) can be obtained by inserting $Wg(z) = A^T \hat{f}_N(Az)$ and $\nabla_z \tilde{H}(z) = \tilde{L}z + A^T f_N(Az)$ into the above equation.

Equation (5.9) yields

$$(5.11) \qquad \|\dot{E}(t)\| \le \|(\nabla_z \tilde{H}(z))^T\| \cdot \|J_{2k}\| \cdot \|A^T\| \cdot \|\hat{f}_N(Az) - f_N(Az)\|.$$

By Lemma 3.2 in [9], there exists a constant $C = \|(P^T \Psi)^{-1}\|$ such that the DEIM approximation error for f_N is bounded by

$$(5.12) \qquad \|\hat{f}_N - f_N\| \le C \|(I - \Psi\Psi^T) f_N\|,$$

where P is the DEIM projection matrix given by (5.3), and Ψ is the collateral POD basis matrix for f_N. Inequalities (5.11) and (5.12) yield

$$(5.13) \qquad \|\dot{E}(t)\| \le \|(\nabla_z \tilde{H}(z))^T\| \cdot \|J_{2k}\| \cdot \|A^T\| \cdot C \|(I - \Psi\Psi^T) f_N\|.$$

Since $A \in Sp(\mathbb{R}^{2n}, 2k)$, we have $\|A\| = 1$. Plugging $\|A^T\| = 1$, $\|J_{2k}\| = 1$, and $\|(\nabla_z \tilde{H}(z))^T\| = \|\nabla_z \tilde{H}(z)\|$ into (5.13), (5.10) follows. $\square$

Strictly speaking, (5.8) is not necessarily Hamiltonian. However, when $\|(I - \Psi\Psi^T) f_N(Az)\| \to 0$, Theorem 5.1 implies that $\|\dot{E}(t)\| \to 0$. In contrast, since the classical DEIM method is devolved to approximate a POD reduced system, the corresponding energy can quickly blow up. Thus, we define *SDEIM* to be the method that applies DEIM to approximate the symplectic Galerkin projection. In particular, for a nonlinear Hamiltonian system $\dot{x} = J_{2n} \nabla_x H(x)$ or $\dot{x} = Kx + J_{2n} f_N(x)$, with $x(0) = x_0$, the SDEIM reduced system is given by (5.8), where the initial condition of the reduced system is given by $z_0 = A^+ x_0$.

Both the cotangent lift (in section 4.1) and the complex SVD (in section 4.2) can be used to construct a symplectic matrix $A \in \mathbb{M}_2(2n, 2k)$ based on an empirical data ensemble. Moreover, (4.16) implies that $\mathbb{M}_2(2n, 2k) \subset V_{2k}(\mathbb{R}^{2n})$. Thus, if we choose Ψ such that

$$(5.14) \qquad A = \Psi \in \mathbb{M}_2(2n, 2k),$$

then $A^T \Psi = I_{2n}$. It follows that $W = (P^T A)^{-1}$. Since $x(t) \in \mathbb{V}$ and $\nabla_x H_2(x) = f_N(x)$, $[x(t); f_N(x(t))]$ is a trajectory in $T^* \mathbb{V} \cong \mathbb{V} \times \mathbb{V}^*$. By assuming $A = \Psi$ in (5.14),

TABLE 1
The POD-Galerkin approach versus the symplectic model reduction approach.

	POD-Galerkin	Symplectic model reduction
Original system	$\dot{x} = f(x)$ with $x \in \mathbb{R}^n$	$\dot{x} = J_{2n}\nabla_x H(x)$ with $x \in \mathbb{R}^{2n}$
Reduced state	Orthogonal projection: $z = \Phi^T x \in \mathbb{R}^k$	Symplectic projection: $z = A^+ x \in \mathbb{R}^{2k}$
Reduced system	Galerkin projection: $\dot{z} = \Phi^T f(\Phi z)$	Symplectic Galerkin projection: $\dot{z} = J_{2k}\nabla_z H(Az)$
Properties of reduced system	No stability guarantee	Energy 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	POD	PSD (a) Cotangent lift (b) Complex SVD (c) NLP
Simplifying nonlinear terms	DEIM: equation (5.5)	SDEIM: equation (5.8)

we actually construct a cotangent lift mapping from $T^*\mathbb{W}$ to $T^*\mathbb{V}$ via a $4n \times 4k$ matrix, $\text{diag}(A, A)$. Using a similar idea from section 4.1, A can be constructed by an extended data ensemble,

$$(5.15) \qquad M_3 := [x(t_1), \ldots, x(t_N), f_N(x(t_1)), \ldots, f_N(x(t_N))],$$

that contains both the state $x(t_i)$ and the nonlinear term $f_N(x(t_i))$.

Regarding the computational complexity of SDEIM, $\tilde{K}$ and W are calculated only once at the beginning. For each step in the online stage, the nonlinear vector term $g(z)$ is evaluated only on selected $2k$ spatial indices of $f_N(Az)$. Thus, the complexity of the SDEIM reduced system (5.8) is also $O(1)$ when k and m' are fixed. Here m' denotes the number of elements of Az that are required to compute the $2k$ spatial indices of $P^T f_N(Az)$.

Table 1 compares the POD-Galerkin approach with the proposed symplectic model reduction approach; it serves as a short summary of sections 3–5.

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 to demonstrate that PSD can deliver stability-preserving reduced systems. Then we study the nonlinear sine-Gordon equation to demonstrate that SDEIM can yield stability-preserving reduced systems with significant speedups.

6.1. Hamiltonian formulation for wave equations. Let $u = u(t, x)$. Consider the one-dimensional semilinear wave equation with constant moving speed c and a nonlinear vector term $g(u)$,

$$(6.1) \qquad u_{tt} = c^2 u_{xx} - g(u),$$

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

$$(6.2) \qquad \dot{q} = \frac{\delta H}{\delta p}, \qquad \dot{p} = -\frac{\delta H}{\delta q},$$

where the Hamiltonian is defined as

$$(6.3) \qquad H(q,p) = \int_0^l dx \left[\frac{1}{2}p^2 + \frac{1}{2}c^2 q_x^2 + G(q) \right], \qquad G'(q) = g(q).$$

A fully resolved model of (6.2) can be constructed by a structure-preserving finite difference discretization [5]. With n equally spaced grid points, the spatial discretized Hamiltonian is given by

$$(6.4) \qquad H_d(y) = \sum_{i=1}^n \Delta x \left[\frac{1}{2}p_i^2 + \frac{c^2(q_{i+1} - q_i)^2}{4\Delta x^2} + \frac{c^2(q_i - q_{i-1})^2}{4\Delta x^2} + G(q_i) \right],$$

where $q_i := u(t, x_i)$, $p_i := u_t(t, x_i)$, $y := [q_1; \ldots; q_n; p_1; \ldots; p_n]$, and $x_i = i\Delta x$. In the limit $\Delta x \to 0$ and $n\Delta x = l$, (6.4) converges to (6.3). Now, the full model is represented by a Hamiltonian ODE system,

$$(6.5) \qquad \frac{dy}{dt} = J_d \nabla_y H_d, \qquad J_d = \frac{J_{2n}}{\Delta x}.$$

Let $D_{xx} \in \mathbb{R}^{n \times n}$ denote the three-point central difference approximation for the spatial derivative ∂_{xx}. We define a Hamiltonian matrix by

$$(6.6) \qquad K = \begin{bmatrix} 0_n & I_n \\ c^2 D_{xx} & 0_n \end{bmatrix}.$$

Then, (6.5) can be written in the form

$$(6.7) \qquad \dot{y} = Ky + J_{2n} f_N(y),$$

where the nonlinear vector term $f_N(y)$ is a vector in $\mathbb{R}^{2n}$ with zeros in the last n elements. Time discretization of (6.7) can be achieved by using the implicit symplectic integrator scheme (2.3).

6.2. Linear wave equation. For our numerical experiments, we first study a linear system with $G(u) = g(u) = 0$ and with periodic boundary conditions. Let $s(x) = 10 \times |x - \frac{1}{2}|$, and let $h(s)$ be a cubic spline function:

$$h(s) = \begin{cases} 1 - \frac{3}{2}s^2 + \frac{3}{4}s^3 & \text{if} \quad 0 \le s \le 1, \\ \frac{1}{4}(2-s)^3 & \text{if} \quad 1 < s \le 2, \\ 0 & \text{if} \quad s > 2. \end{cases}$$

The initial condition is provided by

$$(6.8) \qquad q(0) = [h(s(x_1)); \ldots; h(s(x_n))], \quad p(0) = 0_{n \times 1},$$

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:

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	$c = 0.1$

Regarding reduced systems, we compare all the proposed PSD algorithms (the cotangent lift, complex SVD, and NLP with $r = 100$) with the tangent lift in [18], as well as with POD. Both the Hamiltonian approach (PSD algorithms) and the Lagrangian approach (the tangent lift) are geometric algorithms, as they preserve the Hamiltonian or Lagrangian structures. The criterion for the comparison is the L^2 error between the generalized coordinate $q(t)$ of the benchmark solution and its approximations computed by reduced systems. For the Hamiltonian approach, PSD algorithms are used to construct symplectic matrices to fit the weighted data ensemble (4.24) with $\gamma = \delta t$. For the Lagrangian approach, the basis matrix is constructed from a data ensemble of $q(t)$ in the configuration space. Thus, PSD reduced systems and the reduced system constructed by tangent lift reside on different subspaces of $\mathbb{R}^{2n}$. Nevertheless, because $\dot{q}(t) = p(t)$ holds for the wave equation, symplectic integrators of $[q(t); p(t)]$ can also be used for the time integration of $[q(t); \dot{q}(t)]$ for the fully resolved Lagrangian system. Moreover, the reduced Lagrangian system constructed by the tangent lift also has the form (6.7) if we let $y(t) = [q(t); \dot{q}(t)]$.

Figure 1(a) shows the solution profile at $t = 0$, $t = 2.5$, and $t = 5$. The empirical data ensemble takes 101 snapshots from the benchmark solution trajectory with uniform interval ($\Delta t = 0.5$). We first compare POD with the cotangent lift. When $t = 2.5$, both approaches can obtain good results by taking the first 20 modes; but when $t = 5$, the POD reduced system significantly deviates from the full model.

In Figure 1(b), the blue line represents the singular values of the snapshot matrix M_{x_γ} for POD. Suppose $\{\lambda_1, \ldots, \lambda_k\}$ denote the singular values of the snapshot matrix M_1 (or M_2) of the cotangent lift (or complex SVD). The red (or black) line represents the singular values $\{\lambda_1, \lambda_1, \ldots, \lambda_k, \lambda_k\}$ corresponding to the symplectic basis matrix $A_1 = \mathrm{diag}(\Phi, \Phi)$ (or $A_2 = [\Phi, \Psi; -\Psi, \Phi]$). A fast decay of singular values indicates that a few modes can fit the data with good accuracy. This is a necessary (but not sufficient) condition for a low-dimensional reduced model to approximate the original system with good accuracy. Moreover, we notice that an arbitrary subspace of $\mathbb{R}^{2n}$ can be represented by an orthonormal basis matrix. However, unless this subspace is also symplectic, we cannot represent it by a symplectic basis matrix. Since PSD algorithms can only construct subspaces with the symplectic constraint, both the cotangent lift and complex SVD cannot fit the empirical data as well as POD for the same subspace dimension.

Using more modes, one may expect that both POD and PSD can produce more accurate solutions. However, as Figure 2(a) indicates, the POD reduced system blows up when it has 20 or 40 modes. In addition, the POD reduced system with 40 modes blows up faster than the system with 20 modes. This result verifies that the POD-Galerkin approach can yield unstable reduced systems, even though the original system is stable. By contrast, errors in PSD reduced systems grow slowly in time. Figure 2(b) demonstrates that all the geometric algorithms preserve the system energy E, while the energy of POD reduced systems quickly grows to infinity. Here, E equals the discretized Hamiltonian $H_d(y)$. Let $q_0 = q_n$; (6.4) yields

$$(6.9) \qquad H_d(y) = \frac{\Delta x}{2} \sum_{i=1}^{n} p_i^2 + \frac{c^2}{2\Delta x} \sum_{i=1}^{n} (q_i - q_{i-1})^2.$$

Figure 3 indicates that the L^2 norm of the total error of a POD reduced system is bounded in the interested time domain $[0, 50]$ only when the subspace dimension k is 10 for the cases tested with $10, 20, \ldots, 80$. While reduced systems constructed by geometric algorithms show some numerical error, this error can be systematically
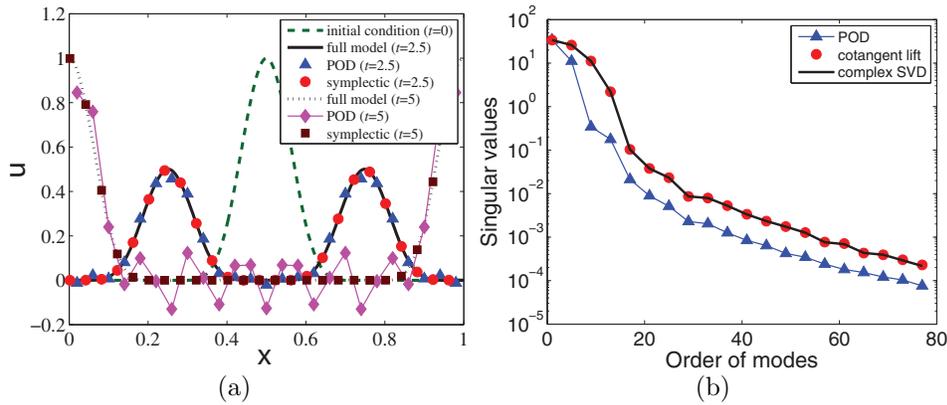
FIG. 1. *(Color online.)* (a) *The solution $u(t,x)$ at $t = 0$, $t = 2.5$, and $t = 5$ of the linear wave equation. We plot the results from the full model, POD, and cotangent lift.* (b) *The first* 80 *singular values corresponding to the first* 80 *POD (or PSD) modes that are used in different reduced systems.*
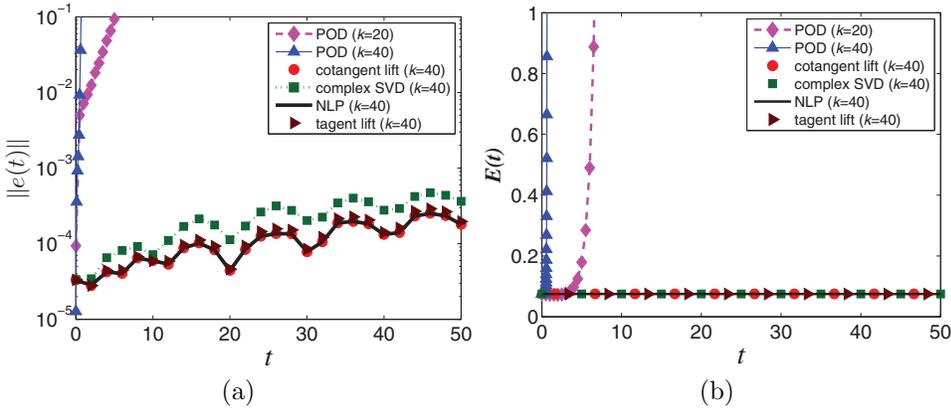

FIG. 2. *(Color online.)* (a) *The evolution of instant L^2 error, $\|e(t)\| := \|\hat{u}(t) - u(t)\|$, between the benchmark solution $u(t)$ and approximating solutions $\hat{u}(t)$ of the linear wave equation.* (b) *The evolution of the energy $E(t)$ of different reduced systems.*
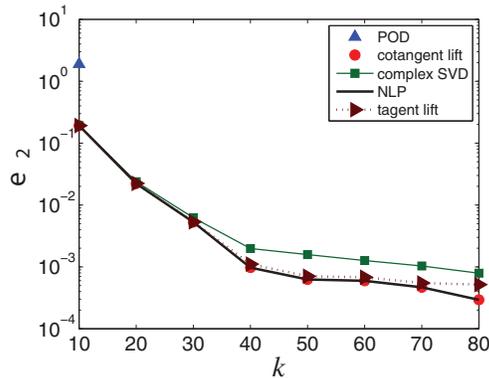

FIG. 3. *(Color online.) The L^2 norm of the total error $\|e\|_2 := \sqrt{\int_0^T \|e(t)\|^2 \mathrm{d}t}$ of different reduced systems for the linear wave equation. For the POD reduced system, we only compute $\|e\|_2$ with 10 modes; when the subspace dimension k is greater than 20, the reduced system blows up in the interested time domain $[0, 50]$ and $\|e\|_2$ becomes infinite.*

reduced by using more modes. In terms of numerical accuracy, the cotangent lift and NLP are slightly better than the tangent lift of [18], while the complex SVD is slightly worse than the tangent lift. NLP yields the most accurate results, but for each k, we observe only a maximum of 0.028% improvement compared with cotangent lift in terms of the relative percentage error.

Stability preservation of symplectic model reduction. To explain our observations mentioned above, we study the stability of the linear wave equation. According to [19], the eigenvalues β_i $(i = 1, \ldots, n)$ of the discretized spatial derivative D_{xx} 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

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

It follows that the eigenvalues of the Hamiltonian matrix K in (6.6) are given by $2n$ pure imaginary numbers $\pm\{\iota\gamma_i\}_{i=1}^n$, where $\gamma_i = c\sqrt{-\beta_i}$, and the corresponding eigenvectors are given by

$$\xi_i := \frac{1}{\sqrt{1 + \gamma_i^2}}\begin{bmatrix} w_i \\ \iota\gamma_i w_i \end{bmatrix}; \qquad \zeta_i := \frac{1}{\sqrt{1 + \gamma_i^2}}\begin{bmatrix} w_i \\ -\iota\gamma_i w_i \end{bmatrix}.$$

Since $\xi_n = \zeta_n = \frac{1}{\sqrt{n}}[1_{n\times 1}; 0_{n\times 1}]$ by the above definition, we can redefine ζ_n to be $\zeta_n = \frac{1}{\sqrt{n}}[0_{n\times 1}; 1_{n\times 1}]$. Thus, we can construct an nonsingular matrix $Q := [\xi_1, \zeta_1, \ldots, \xi_n, \zeta_n]$ such that K is transformed to a 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 ζ_n, the projection coefficient of initial condition (6.8) onto this mode vanishes, i.e., $\zeta_n^T y_0 = 0$. Thus, the original system evolves on a stable subspace of $\mathbb{R}^{2n}$.

Next, we consider the reduced system constructed by the symplectic projection. By (6.9), we have $H_d(y) \geq 0$, and the equality holds if and only if y is parallel to ξ_n. If $\xi_n \notin \text{Range}(A)$, the origin of $\mathbb{R}^{2n}$ is the only solution that satisfies $H_d(y) = 0$ for all $y \in \text{Range}(A)$. As a consequence, the original is the strict minimum of H_d in $\text{Range}(A)$. In our numerical simulations, we do observe that $A^+\xi_n \neq 0$, where A is constructed by the cotangent lift, complex SVD, or NLP. Then, Theorem 3.10 implies that the origin is a stable equilibrium for the reduced Hamiltonian system.

Instability of POD-Galerkin. Since POD does not preserve the system energy, there are no mechanisms similar to the Hamiltonian and Lagrangian approaches that limit the solution trajectory in a bounded region. As a result, the reduced system may blow up with time evolution. To corroborate this claim, let λ_* denote the eigenvalue of $\Phi^T K\Phi$ with the maximal real part and let ξ_* denote the corresponding eigenvector with unit length. Then, $a_* = \xi_*^T y_0$ gives the projection coefficient of y_0 onto ξ_*. Table 2 indicates that for each subspace dimension k, a POD reduced system has $\text{Re}(\lambda_*) > 0$ and $a_* \neq 0$. Since the solution of a linear system has an exponential term $a_* \exp(\lambda_* t)\xi_*$, the POD reduced system is always unstable for long-time integration.

TABLE 2

The eigenvalue λ_* and coefficient a_* of a POD unstable mode $a_* \exp(\lambda_* t)\xi_*$ for different subspace dimension k.

k	10	20	30	40	50	60	70	80
λ_*	0.0338	0.659	13.74	14.39	14.50	5.33	10.42	$13.05 + 5.09\,\iota$
a_*	0.929	0.0184	0.0263	0.0498	-0.0200	0.0718	7.55e-3	-0.0068 - 0.0142ι

Assume that Φ_{k_1} and Φ_{k_2} respectively contain the first k_1 and k_2 dominant modes. If $k_1 < k_2$, then $\Phi_{k_1}^T K \Phi_{k_1}$ is a submatrix of $\Phi_{k_2}^T K \Phi_{k_2}$, and $\|\Phi_{k_1}^T K \Phi_{k_1}\| \le \|\Phi_{k_2}^T K \Phi_{k_2}\|$ holds. As $\mathrm{Re}(\lambda_*) \le |\lambda_*| \le \|\Phi^T K \Phi\|$, the matrix norm of $\Phi^T K \Phi$ provides an upper bound to $\mathrm{Re}(\lambda_*)$. Thus, the upper bound to $\mathrm{Re}(\lambda_*)$ is a monotonically increasing function of the subspace dimension k. Table 2 also shows that $\mathrm{Re}(\lambda_*)$ with 40 modes is much larger than $\mathrm{Re}(\lambda_*)$ with 20 modes, which explains why the POD reduced system with 40 modes blows up faster than the system with 20 modes in Figure 2. Although for $k = 10$, the POD reduced system can produce accurate solution for a short time domain $[0, 2.5]$, we can still observe that this system blows up for a large enough integration time, say, $t > 10$.

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, which was first studied in the 1970s, appears in a number of physical applications, including relativistic field theory, Josephson junctions, and mechanical transmission lines [31]. One can show that the sine-Gordon equation admits a localized solitary wave solution,

$$(6.10) \qquad u(t,x) = 4\arctan\left[\exp\left(\pm \tfrac{x-x_0-vt}{\sqrt{1-v^2}}\right)\right],$$

which travels with the speed $|v| < 1$. The $\pm$ signs correspond to localized solutions which are called *kink* and *antikink*, respectively [31].

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

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	$v = 0.2$

The L^2 error for the state variable $y(t)$ is studied for the full model and reduced models constructed by POD, cotangent lift, complex SVD, DEIM, and SDEIM. All the basis matrices are constructed to fit the data ensemble (4.24) with $\gamma = 1$. For SDEIM, the cotangent lift is used to construct a symplectic basis matrix $A_1 = \Psi = \mathrm{diag}(\Phi, \Phi)$ to fit both the state variable $y(t)$ and the nonlinear vector $f_N(y(t))$, where Φ is a POD basis matrix for the extended snapshot matrix that contains $q(t)$, $p(t)$, and $f_N(q(t))$ in its column vectors.

Figure 4(a) shows the *kink* solution profile at $t = 0$, $t = 25$, and $t = 75$. The data ensemble takes 1201 snapshots from the solution trajectory, solved by the full model with uniform interval ($\Delta t = 0.125$). We first compare POD with the cotangent lift. For short-time integration, both approaches can obtain very accurate results by taking the first 60 modes. In Figure 4(b), we study the singular values corresponding
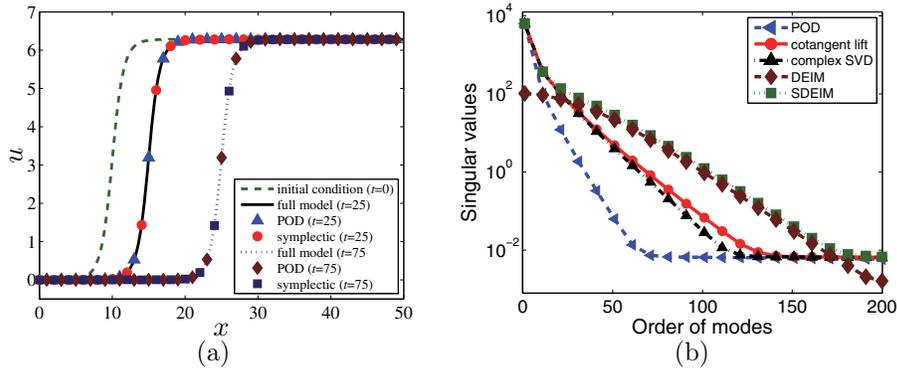
FIG. 4. *(Color online.)* (a) *The solution $u(t, x)$ at $t = 0$, $t = 25$, and $t = 75$ of the sine-Gordon equation. We plot the results from the full model, POD, and cotangent lift.* (b) *The first 200 singular values corresponding to the first 200 POD (or DEIM, PSD, SDEIM) modes that are used in different reduced systems.*
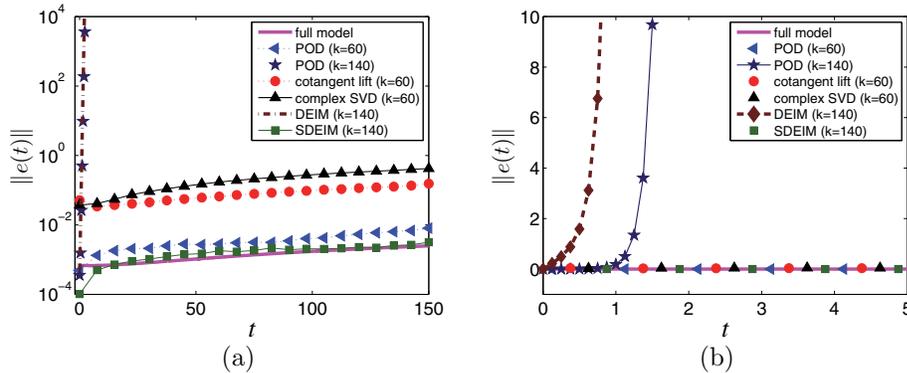

FIG. 5. *(Color online.)* (a) *The evolution of instant L^2 error, $\|e(t)\| := \|\hat{y}(t) - y(t)\|$, between the analytic solution $y(t)$ in the phase space and approximating solutions $\hat{y}(t)$ of the sine-Gordon equation for $t \in [0, 150]$.* (b) *The instant L^2 error $\|e(t)\|$ for $t \in [0, 5]$.*

to the POD basis matrix, the PSD basis matrices constructed by the cotangent lift and complex SVD, the nonlinear term basis matrix for DEIM, and the symplectic matrix for SDEIM. This figure demonstrates that POD is better to fit empirical state variables than the cotangent lift and complex SVD, while DEIM is better to fit empirical nonlinear vectors than SDEIM.

Figure 5 illustrates that all symplectic schemes (including the cotangent lift, complex SVD, and SDEIM) yield low computational errors with appropriate subspace dimension, while nonsymplectic schemes (including the POD and DEIM) can yield unbounded numerical error with 140 modes. In Figure 6, all symplectic schemes can effectively preserve the system energy E. By contrast, both POD and DEIM reduced systems can achieve infinite energy with 140 modes. Here, E equals the discretized Hamiltonian $H_d(y)$. With $G(u) = 1 - \cos(u)$ and Dirichlet boundary conditions, (6.4) gives

$$(6.11) \quad H_d(y) = \frac{\Delta x}{2} \sum_{i=1}^{n} p_i^2 + \Delta x \sum_{i=1}^{n} [1 - \cos(q_i)] + \frac{q_1^2}{4\Delta x} + \frac{1}{2\Delta x} \sum_{i=2}^{n} (q_i - q_{i-1})^2 + \frac{(q_n - 2\pi)^2}{4\Delta x}.$$

Figure 7(a) indicates that by using more modes, all symplectic reduced models can obtain better accuracy and finally converge to the full model. By contrast, POD
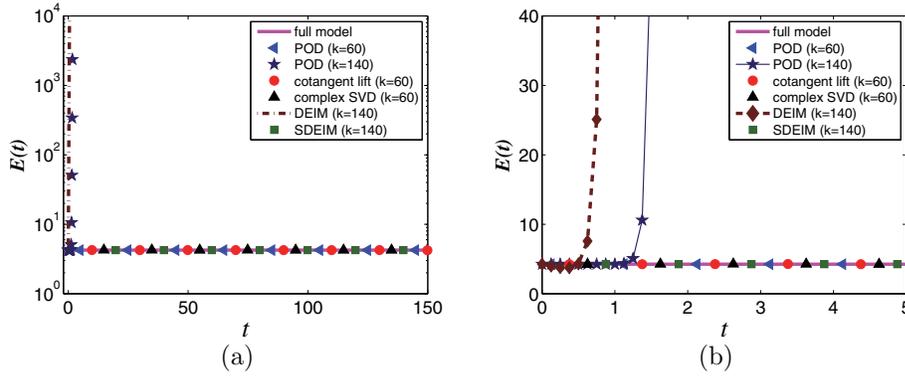
FIG. 6. *(Color online.)* (a) *The evolution of the system energy $E(t)$ of the sine-Gordon equation for $t \in [0, 150]$.* (b) *The evolution of the system energy $E(t)$ for $t \in [0, 5]$.*
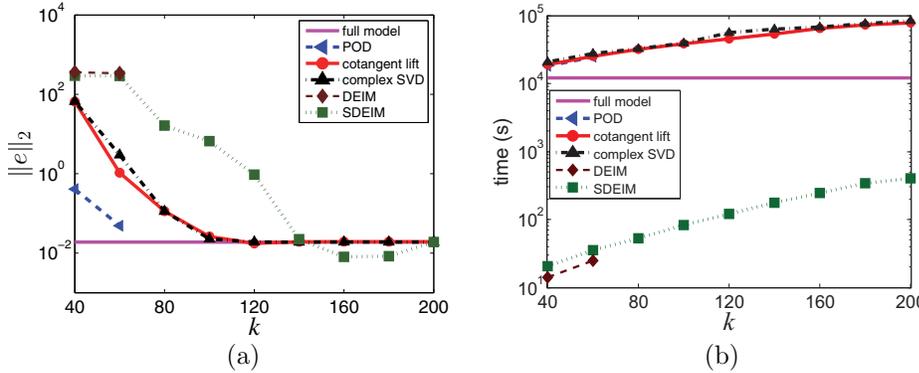

FIG. 7. *(Color online.)* (a) *The L^2 norm of the total error $\|e\|_2 := \sqrt{\int_0^T \|e(t)\|^2 dt}$ of the full model and different reduced models for the sine-Gordon equation. For the POD and DEIM reduced systems, we only compute $\|e\|_2$ for $k = 40$ and $k = 60$; when $k \geq 80$, the reduced systems blow up in the interested time domain $[0, 150]$ and $\|e(t)\|_2$ becomes infinite.* (b) *The running time of different model reduction techniques with different k. All the data come from the average value of ten independent runs.*

and DEIM can yield unbounded reduced systems in the interested time domain $[0, 150]$ when the subspace dimension k is greater than 80 for the cases tested with $40, 60, \ldots, 200$. Furthermore, by the analysis in section 5, we know that a direct use of POD or PSD is not able to obtain any speedups for the sine-Gordon equation, since it contains a nonlinear vector term $\sin(u)$. Numerical results in Figure 7(b) also verify this point. Especially, the running time for POD, the cotangent lift, and complex SVD is even larger than the running time for the full model. On the contrary, both the DEIM and SDEIM approximations can significantly improve the computational efficiency and reduce the running time of POD or PSD by three orders of magnitude.

The boundedness of Hamiltonian systems can be derived by their energy conservation property. If $y(t)$ denotes the solution trajectory, we have $H_d(y(t)) = E$ for a constant E. Since each term on the right-hand side of (6.11) 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 $y \in \mathbb{R}^{2n}$, as long as $\|y\| = M$, we have $H_d(y) > E$. Therefore, by Theorem 3.9 both the original system and reduced systems constructed by the symplectic projection are bounded for all t.

7. Conclusion. In this paper, we proposed a symplectic model reduction technique for the reduced-order modeling of large-scale Hamiltonian systems. We first defined the symplectic projection, which can yield reduced systems that remain Hamiltonian. Several PSD algorithms, such as the cotangent lift, complex SVD, and NLP, 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 stability. Thus, the proposed technique is very suited for long-time integration, especially when the original systems are conservative and do not have any natural dissipative mechanism to stabilize them. Since the symplectic projection can only speed up linear and quadratic problems, the PSD was also combined with DEIM, effectively reducing 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. We demonstrated the capability of the symplectic model reduction to solve a large-scale system with high accuracy, good efficiency, and stability preservation via linear and nonlinear wave equations.

REFERENCES

[1] M. Amabili, A. Sarkar, and M. P. Païdoussis, *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 methods for Hamiltonian PDEs*, J. Phys. A, 39 (2006), pp. 5287–5320.

[6] T. Bui-Thanh, M. Damodaran, and K. Willcox, *Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition*, AIAA J., 42 (2004), pp. 1505–1516.

[7] K. Carlberg, C. Bou-Mosleh, and C. Farhat, *Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations*, Internat. J. Numer. Methods Engrg., 86 (2011), pp. 155–181.

[8] K. Carlberg, R. Tuminaro, and P. Boggs, *Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics*, SIAM J. Sci. Comput., 37 (2014), pp. B153–B184.

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

[10] 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.

[11] H. Eves, *Elementary Matrix Theory*, Dover, New York, 1980.

[12] D. Galbally, K. Fidkowski, K. Willcox, and O. Ghattas, *Non-linear model reduction for uncertainty quantification in large-scale inverse problems*, Internat. J. Numer. Methods Engrg., 81 (2010), pp. 1581–1608.

[13] 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.

[14] S. Gugercin, R. V. Polyuga, C. Beattie, and A. J. van der Schaft, *Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems*, Automatica, 48 (2012), pp. 1963–1974.

[15] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Math. 31, Springer-Verlag, Berlin, 2006.

[16] C. Hartmann, V. -M. Vulcanov, and C. Schütte, *Balanced truncation of linear second-order systems: A Hamiltonian approach*, SIAM J. Multiscale Model. Simul., 8 (2010), pp. 1348–1367.

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

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

[19] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, SIAM, Philadelphia, 2007.

[20] J. E. Marsden and T. Ratiu, *Introduction to Mechanics and Symmetry*, 2nd ed., Texts in Appl. Math. 17, Springer-Verlag, New York, 2003.

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

[22] J. D. Meiss, *Differential Dynamical Systems*, SIAM, Philadelphia, 2007.

[23] K. R. Meyer, G. R. Hall, and D. Offin, *Introduction to Hamiltonian dynamical systems and the N-body problem*, 2nd ed., Appl. Math. Sci. 90, Springer-Verlag, New York, 2009.

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

[25] P. A. Parrilo, S. Lall, F. Paganini, G. C. Verghese, B. C. Lesieutre, and J. E. Marsden, *Model reduction for analysis of cascading failures in power systems*, in Proceedings of the American Control Conference, Vol. 6, San Diego, CA, 1999, pp. 4208–4212.

[26] S. Prajna, *POD model reduction with stability guarantee*, in Proceedings of the 42nd IEEE Conference on Decision and Control, Vol. 5, Maui, HI, 2003, pp. 5254–5258.

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

[28] M. Rewieński 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.

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

[30] R. V. Polyuga and A. J. van der Schaft, *Structure preserving model reduction of port-Hamiltonian systems by moment matching at infinity*, Automatica, 46 (2010), pp. 665–672.

[31] G. B. Whitham, *Linear and Nonlinear Waves*, Wiley-Interscience, New York, 1999.

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