

Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 26/27

Matthias Ehrhardt and Achraf Zinihi

Convex Structure-Preserving EDMD: Passive Koopman Matrices as Q-Contractions

November 20, 2026

http://www.imacm.uni-wuppertal.de

Convex Structure-Preserving EDMD: Passive Koopman Matrices as Q-Contractions

Matthias Ehrhardt[a,*], Achraf Zinihi[a]

[a] *Applied and Computational Mathematics, University of Wuppertal, Gaußstrasse 20, Wuppertal, 42119, Germany*

Abstract

Extended dynamic mode decomposition (EDMD) approximates the Koopman operator using a matrix based on a finite set of observables. However, the least-squares matrix of a port-Hamiltonian (pH) system is generally neither passive nor invariant-preserving. We demonstrate that the implicit midpoint discretization of a lifted pH model precisely preserves energy balance, passivity, and all linear Casimirs for every step size. For a fixed storage metric Q, the resulting Koopman matrices are the Cayley transforms of $(J - R)Q$, where J is skew-symmetric and R is positive semidefinite. This set does not depend on the step size and is the closure of all Q-contractions. This set is LMI-representable; therefore, structure-preserving EDMD, including ports through a discrete-time Kalman-Yakubovich-Popov condition and Casimirs through affine constraints, is a semidefinite program. The constrained estimate is a projection of the unconstrained estimate in the data-weighted norm and cannot increase estimation error when the population target is pH-compatible. For every Q-contraction, a certified bound on the lifted prediction error grows linearly at most in time, whereas it grows geometrically for unconstrained models with Q-gain above one. Benchmarks on a damped pendulum and a damped rigid body demonstrate that constrained models never generate energy and that imposed Casimirs are maintained with solver accuracy (approximately 10^{-8}). In contrast, EDMD may diverge or lose the Casimir with noisy data. The benchmarks also demonstrate that when Q is the physical energy, the constrained models are less accurate than EDMD and can exhibit significant structural bias.

Keywords: Koopman operator, extended dynamic mode decomposition, port-Hamiltonian systems, passivity, Cayley transform, linear matrix inequalities, Casimir functions

2020 Mathematics Subject Classification: 93B30, 93C55, 37M99, 90C22.

1. Introduction

Extended dynamic mode decomposition (EDMD) [1, 2] approximates the Koopman operator [3–6] of a nonlinear system with a matrix that acts on a finite dictionary of observables. It has become a standard tool for prediction and predictive control [7–9]. However, for port-Hamiltonian (pH) systems [10, 11] the unconstrained least-squares matrix is structurally agnostic: it need not be passive with respect to any storage function, nor does it need to conserve the Casimirs of the system. Long-horizon simulations often exhibit spurious energy drift.

Structure-preserving fits are available for linear pH systems in state coordinates [12–14]., The operator-theoretic structure of the Koopman generator of nonlinear pH systems has also been clarified [15]. In this letter, we address the discrete-time identification problem in lifted coordinates and demonstrate that it is a convex program for a fixed storage metric.

*Corresponding author

Email addresses: ehrhardt@uni-wuppertal.de (Matthias Ehrhardt), achrafzinihi99@gmail.com (Achraf Zinihi)

The main contributions of this letter are as follows: First, we demonstrate that the implicit midpoint discretization of a lifted pH model exactly preserves the energy balance, passivity, and linear Casimirs for all step sizes. Second, we show that the resulting Koopman matrices with the storage function $\frac{1}{2}z^\top Qz$ are precisely the Q-contractions, independent of the step size Δt. Consequently, structure-preserving EDMD with ports and Casimirs is a semidefinite program (SDP). The solution to this program is a projection of the unconstrained estimate in the data-weighted norm. Therefore, the constraint cannot increase the estimation error when the population target is pH-compatible. Furthermore, the lifted prediction error of the model grows at most linearly with time. Three benchmarks quantify the benefits and costs of the constraints.

The Cayley (bilinear, Tustin) transform and the discrete-time Kalman-Yakubovich-Popov (KYP) Lemma are classical. Discrete-time pH systems based on symplectic integration and collocation have been studied in references [16–20]. This letter contributes their combination in the lifted Koopman setting, yielding an identification method that is exactly passive and Casimir-preserving while remaining convex. A systematic choice of the storage metric, design of admissible dictionaries, and sharper error bounds are topics for future research.

Three recent works are the most similar to ours. Lortie et al. [21] and Mamakoukas et al. [22] constrain the EDMD matrix to be stable; however, stability alone does not prevent transient energy growth. Singh et al. [23] learn a lifting with a neural network together with a pH generator, discretize it using a Cayley (midpoint) step, and prove zero-input contractivity. Thus, the midpoint step itself is not novel. This letter adds an exact description of the entire passive set independent of Δt, a direct convex fit of A_d from snapshot pairs with ports and Casimirs, a projection property, and a quantitative account of accuracy costs.

2. The Problem Setting

We consider input-state-output pH systems of the form

$$\dot{x} = \big[J(x) - R(x)\big]\nabla H(x) + G(x)u, \quad \text{with} \quad y = G(x)^\top \nabla H(x), \tag{1}$$

on $\mathcal{X} \subseteq \mathbb{R}^n$, with a *Hamiltonian* $H \in C^2(\mathcal{X})$ that is bounded below, with a skew-symmetric interconnection matrix $J(x) = -J(x)^\top$, and a symmetric positive semidefinite dissipation matrix $R(x) = R(x)^\top \succeq 0$. Along solutions of (1), we have the *dissipation inequality*

$$\frac{d}{dt}H(x(t)) = -\nabla H^\top R \nabla H + u^\top y \leq u^\top y, \tag{2}$$

i.e., the pH system (1) is passive with storage function H [24]. We recall that a function C with the property $J\nabla C \equiv 0$ is a Casimir [25].

Next, let $\Psi = (\psi_1, \ldots, \psi_N)^\top$ be a *dictionary of observables* and $z = \Psi(x)$ the *lifted state*. Given snapshot pairs (x_k, x_k^+), $k = 1, \ldots, K$, where x_k^+ denotes the state one sampling period Δt after x_k, standard EDMD computes

$$\widehat{K} = \arg\min_{K \in \mathbb{R}^{N\times N}} \big\| Z_Y - K Z_X \big\|_F^2, \tag{3}$$

where

$$Z_X = \big[\Psi(x_1), \ldots, \Psi(x_K)\big], \quad \text{and} \quad Z_Y = \big[\Psi(x_1^+), \ldots, \Psi(x_K^+)\big].$$

For the continuous-time model class, we use the *lifted pH system*

$$\dot{z} = \Big(J_\mathrm{K} - R_\mathrm{K} + \sum_{j=1}^m u_j S_j\Big)Qz + Bu, \quad \text{with} \quad y_{\mathrm{K},j} = (Qz)^\top S_j (Qz) + \big(B^\top Qz\big)_j, \tag{4}$$

with $Q = Q^\top \succ 0$, $J_\mathrm{K} = -J_\mathrm{K}^\top$, $R_\mathrm{K} = R_\mathrm{K}^\top \succeq 0$, arbitrary $S_j \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times m}$, and lifted storage function $H_\mathrm{K}(z) = \frac{1}{2}z^\top Qz$. Direct computation using $Qz^T JK(Qz) = 0$ yields the following (cf. (2))

$$\frac{d}{dt}H_\mathrm{K} = -(Qz)^\top R_\mathrm{K}(Qz) + y_\mathrm{K}^\top u \leq y_\mathrm{K}^\top u,$$

for all parameter values. Thus, the lifted pH system (4) is passive by construction. The bilinear terms S_j are necessary for an exact representation of the energy balance of control-affine systems. Throughout this paper, the storage metric Q is given. Section 6 discusses how it can be chosen.

3. Exact Discrete-Time Structure Preservation

Continuous-time passivity (2) is not inherited by arbitrary discretizations. For example, the energy increment of the explicit Euler method contains the nonnegative remainder $\frac{\Delta t^2}{2} f^\top Q f$, where f is the right-hand side of (4), which can destroy passivity. Therefore, we work with the implicit midpoint rule.

Definition 3.1 (Discrete-time lifted pH system)**.** Given the parameters of (4) and $\Delta t > 0$, the *implicit midpoint scheme* reads

$$z_{k+1} = z_k + \Delta t\Big[\Big(J_{\mathrm{K}} - R_{\mathrm{K}} + \sum_{j=1}^{m} u_{k,j} S_j\Big) Q z_{k+1/2} + B u_k\Big], \quad \text{where} \quad z_{k+1/2} := \tfrac{1}{2}(z_k + z_{k+1}), \tag{5}$$

and the outputs are $y_{\mathrm{K},k,j} := (Q z_{k+1/2})^\top S_j (Q z_{k+1/2}) + (B^\top Q z_{k+1/2})_j$.

The scheme requires solving one linear system per step, with the following matrix $I - \frac{\Delta t}{2}(J_{\mathrm{K}} - R_{\mathrm{K}} + \sum_j u_{k,j} S_j) Q$. For $S_j = 0$, this matrix is invertible for all $\Delta t > 0$ (Proposition 4.1). For $S_j \neq 0$, unique solvability depends on the values of both the time step and the input amplitude.

Theorem 3.2 (Exact discrete structure preservation)**.** *Let $\Delta t > 0$, and suppose that (5) is uniquely solvable at every step. Then the scheme (5) has the following properties:*

(*i*) exact energy balance *(as a discrete analogue to the dissipation inequality (2)):*

$$H_{\mathrm{K}}(z_{k+1}) - H_{\mathrm{K}}(z_k) = -\Delta t\, (Q z_{k+1/2})^\top R_{\mathrm{K}} (Q z_{k+1/2}) + \Delta t\, u_k^\top y_{\mathrm{K},k} \leq \Delta t\, u_k^\top y_{\mathrm{K},k}. \tag{6}$$

Hence, discrete-time passivity with storage function H_{K};

(*ii*) exact Casimir preservation: *if $c \in \ker J_{\mathrm{K}} \cap \ker R_{\mathrm{K}}$ and, in the forced case, $c^\top S_j = 0$ and $c^\top B = 0$, then $c^\top z_{k+1} = c^\top z_k$;*

(*iii*) second-order accuracy: *the scheme is a second-order approximation of the flow of (4).*

Proof. (*i*) Since $Q = Q^\top$, we have the energy balance $H_{\mathrm{K}}(z_{k+1}) - H_{\mathrm{K}}(z_k) = (z_{k+1} - z_k)^\top Q z_{k+1/2}$. Substituting the implicit midpoint scheme (5) and using $(Q z_{k+1/2})^\top J_{\mathrm{K}} (Q z_{k+1/2}) = 0$ gives (6).

(*ii*) Multiplying the implicit midpoint scheme (5) by $c^\top$ and using $c^\top J_{\mathrm{K}} = c^\top R_{\mathrm{K}} = 0$, which follows from $J_{\mathrm{K}} c = R_{\mathrm{K}} c = 0$ by skew symmetry and symmetry, respectively, together with $c^\top S_j = 0$ and $c^\top B = 0$, gives $c^\top (z_{k+1} - z_k) = 0$.

(*iii*) This is the classical order of accuracy of the implicit midpoint rule. □

Theorem 3.2 shows that the continuous pH structure carries over to a discrete energy inequality *exactly*, at the cost of one linear solve per step, and that the Casimirs are carried along. For a nonlinear lifted Hamiltonian, one replaces $Q z_{k+1/2}$ with a discrete gradient $\overline{\nabla} H_{\mathrm{K}}(z_k, z_{k+1})$, cf. [18, 20, 26], and the above mentioned properties (*i*)–(*ii*) remain valid verbatim.

4. Passive Koopman Matrices are Q-Contractions

The structure-preserving version of (3) restricts K to the set of Koopman matrices that admit a pH realization. The following result identifies this set.

Proposition 4.1 (Cayley correspondence). *Let $Q \succ 0$ and $\Delta t > 0$. For $\mathcal{A} \in \mathbb{R}^{N\times N}$, the following are equivalent:*

(a) $\mathcal{A} = (J_{\mathrm{K}} - R_{\mathrm{K}})Q$ *for some* $J_{\mathrm{K}} = -J_{\mathrm{K}}^\top$ *and* $R_{\mathrm{K}} \succeq 0$;

(b) $Q\mathcal{A} + \mathcal{A}^\top Q \preceq 0$.

In this case, $I - \frac{\Delta t}{2}\mathcal{A}$ is invertible, and the Cayley transform

$$A_d = \mathrm{Cay}_{\Delta t}(\mathcal{A}) := \Big(I - \frac{\Delta t}{2}\mathcal{A}\Big)^{-1}\Big(I + \frac{\Delta t}{2}\mathcal{A}\Big)$$

*satisfies the Q-*contraction property *$A_d^\top Q A_d \preceq Q$ and inherits the linear invariants: $c^\top \mathcal{A} = 0$ implies $c^\top A_d = c^\top$. Conversely, every A_d with $A_d^\top Q A_d \preceq Q$ and $-1 \notin \sigma(A_d)$ is the Cayley transform of a unique such $\mathcal{A}$.*

Proof. (a) $\Rightarrow$ (b): $Q\mathcal{A} + \mathcal{A}^\top Q = -2QR_{\mathrm{K}}Q \preceq 0$.

(b) $\Rightarrow$ (a): Take $J_{\mathrm{K}} = \frac{1}{2}(\mathcal{A}Q^{-1} - Q^{-1}\mathcal{A}^\top)$ and $R_{\mathrm{K}} = -\frac{1}{2}(\mathcal{A}Q^{-1} + Q^{-1}\mathcal{A}^\top)$. $R_{\mathrm{K}} \succeq 0$ is exactly (b) after congruence with Q^{-1}. By (b), the spectrum of $\mathcal{A}$ lies in the closed left half-plane, so $2/\Delta t$ is not an eigenvalue and $I - \frac{\Delta t}{2}\mathcal{A}$ is invertible. With $W = I - \frac{\Delta t}{2}\mathcal{A}$ and $V = I + \frac{\Delta t}{2}\mathcal{A}$, which commute,

$$A_d^\top Q A_d - Q = W^{-\top}\big(V^\top Q V - W^\top Q W\big)W^{-1}, \quad \text{and} \quad V^\top Q V - W^\top Q W = \Delta t\,(\mathcal{A}^\top Q + Q\mathcal{A}) \preceq 0.$$

If $\mathcal{A}^\top c = 0$, then $W^\top c = c$, hence $W^{-\top}c = c$ and $A_d^\top c = V^\top W^{-\top} c = V^\top c = c$. For the converse, the inverse transform is $\mathcal{A} = \frac{2}{\Delta t}(A_d + I)^{-1}(A_d - I)$, and the computation above, read backwards, shows that $A_d^\top Q A_d \preceq Q$ implies (b). □

We write $\|v\|_Q := (v^\top Q v)^{1/2}$, and use the same symbol for the induced matrix norm, so that $A_d^\top Q A_d \preceq Q$ is equivalent to $\|A_d\|_Q \le 1$. We denote the set of Q-contractions by $\mathcal{S}_{\mathrm{pH}}(Q) := \{A_d \in \mathbb{R}^{N\times N} \colon A_d^\top Q A_d \preceq Q\}$. The subset $\{A_d \in \mathcal{S}_{\mathrm{pH}}(Q) \colon -1 \notin \sigma(A_d)\}$ in the converse part of Proposition 4.1 does not involve Δt; only the generator $\mathcal{A}$ attached to a given A_d does. Furthermore, if $A_d \in \mathcal{S}_{\mathrm{pH}}(Q)$, then $A_\varepsilon = (1-\varepsilon)A_d$ satisfies $\|A_\varepsilon\|_Q \le 1 - \varepsilon$, so $-1 \notin \sigma(A_\varepsilon)$ and $A_\varepsilon \to A_d$ as $\varepsilon \to 0$. Therefore, the Cayley images of Q-dissipative generators are dense in the closed set $\mathcal{S}_{\mathrm{pH}}(Q)$. Therefore, searching over all Q-contractions loses nothing, up to closure, compared with searching over Cayley-discretized pH models, and no derivative data or generator parametrization is needed.

Corollary 4.2 (Convex structure-preserving EDMD). *For a fixed $Q \succ 0$, the set $\mathcal{S}_{\mathrm{pH}}(Q)$ is closed, convex, and representable by a linear matrix inequality (LMI)*

$$\begin{bmatrix} Q & A_d^\top Q \\ Q A_d & Q \end{bmatrix} \succeq 0, \tag{7}$$

so that $\min_{A_d \in \mathcal{S}_{\mathrm{pH}}(Q)} \|Z_Y - A_d Z_X\|_F^2$ is a convex SDP. With ports, the passivity of (A_d, B_d, C_d, D_d) with storage function $\frac{1}{2}z^\top Q z$ and supply rate $u^\top y$ is equivalent to the discrete-time KYP condition

$$\begin{bmatrix} Q & C_d^\top & A_d^\top \\ C_d & D_d + D_d^\top & B_d^\top \\ A_d & B_d & Q^{-1} \end{bmatrix} \succeq 0, \tag{8}$$

which is again an LMI in (A_d, B_d, C_d, D_d) for a fixed Q. Imposing the linear Casimir $C_0^\top z$ adds the affine constraints $C_0^\top A_d = C_0^\top$ and $C_0^\top B_d = 0$, which preserve convexity.

Proof. Both LMIs follow from a Schur complement applied to

$$[A_d\ B_d]^\top Q[A_d\ B_d] - \begin{bmatrix} Q & C_d^\top \\ C_d & D_d + D_d^\top \end{bmatrix} \preceq 0,$$

which is the dissipation inequality $\frac{1}{2}z_{k+1}^\top Q z_{k+1} - \frac{1}{2}z_k^\top Q z_k \le u_k^\top y_k$ written as a quadratic form in (z_k, u_k). For (7), set $B_d = 0$, $C_d = 0$, and $D_d = 0$, delete the resulting zero block row and column, and multiply the last block row and column by Q. □

The matrices (A_d, B_d) above describe models that are linear in the input. The bilinear terms S_j in (4) cause the midpoint map to be nonlinear in (z_k, u_k), so Corollary 4.2 covers the case $S_j = 0$, which is also the case used in Section 7. The Casimir constraints never render the program infeasible. For C_0 with full column rank, $A_d = Q^{-1}C_0(C_0^\top Q^{-1}C_0)^{-1}C_0^\top$, $B_d = 0$ is the Q-orthogonal projector onto range$(Q^{-1}C_0)$. Thus, it is a Q-contraction, that satisfies both constraints.

Corollary 4.2 is the algorithmic core of the approach. For a fixed storage metric, structure-preserving EDMD is a convex program, requiring only snapshot pairs, not derivative data. Its continuous-time counterpart is the LMI $Q\mathcal{A} + \mathcal{A}^\top Q \preceq 0$ of Proposition 4.1(*b*). The nonconvexity of the identification problem lies entirely in jointly estimating Q and the remaining parameters. This is why Q is fixed before the regression. After solving the SDP, a pH realization $(J_\mathrm{K}, R_\mathrm{K})$ with $\mathrm{Cay}_{\Delta t}((J_\mathrm{K} - R_\mathrm{K})Q) = A_d$ is obtained using the inverse Cayley transform and the formulas in the proof of Proposition 4.1, provided that $-1 \notin \sigma(A_d)$.

5. Projection Property and Error Growth

Given $z_k = \Psi(x_k)$ and $z_k^+ = \Psi(x_k^+)$, we can write the regression equation (3), possibly augmented by an input matrix, as a linear least-squares problem in a parameter vector $\xi \in \mathbb{R}^p$: for $\xi = \mathrm{vec}(A_d)$, we have $A_d z_k = \Phi(z_k)\xi$ with $\Phi(z_k) = z_k^\top \otimes I_N$.

Let $\mathcal{C} \subseteq \mathbb{R}^p$ be the closed convex constraint set of Corollary 4.2, and define the *population target*, cf. [27, Section 5.3.2], (also called $L_2(\nu)$-projection in the foundational paper [2]):

$$\xi^\star = \arg\min_{\xi\in\mathbb{R}^p} \mathbb{E}\big\|z^+ - \Phi(z)\xi\big\|^2,$$

where the expectation is taken with respect to the sampling distribution of the snapshot pairs.

Define the *empirical Gram matrix* $\widehat{G} := \frac{1}{K}\sum_{k=1}^K \Phi(z_k)^\top\Phi(z_k) = \big(\frac{1}{K}Z_X Z_X^\top\big) \otimes I_N$, which is assumed to be positive definite. Define the $\widehat{G}$-induced norm as follows: $\|\xi\|_{\widehat{G}} := (\xi^\top \widehat{G}\xi)^{1/2}$.

Proposition 5.1 (Nonexpansiveness of the structural constraint)**.** *Let $\mathcal{C}$ be a closed, convex set. Let $\widehat{\xi}$ be the unconstrained least-squares estimate, $\widehat{\xi}_\mathcal{C}$ the constrained one. Let the the population target be denoted by $\xi^\star$ Then $\widehat{\xi}_\mathcal{C} = \Pi_\mathcal{C}(\widehat{\xi})$ is the metric projection of $\widehat{\xi}$ onto $\mathcal{C}$ in the norm $\|\cdot\|_{\widehat{G}}$, and*

$$\|\widehat{\xi}_\mathcal{C} - \xi^\star\|_{\widehat{G}} \le \|\widehat{\xi} - \xi^\star\|_{\widehat{G}} + \mathrm{dist}_{\widehat{G}}(\xi^\star, \mathcal{C}).$$

In particular, if $\xi^\star \in \mathcal{C}$, that is, if the population target is itself pH-compatible, then the constrained estimator is never worse than the unconstrained one: $\|\widehat{\xi}_\mathcal{C} - \xi^\star\|_{\widehat{G}} \le \|\widehat{\xi} - \xi^\star\|_{\widehat{G}}$.

Proof. By the normal equations, the empirical loss satisfies

$$\frac{1}{K}\sum_{k=1}^K \big\|z_k^+ - \Phi(z_k)\xi\big\|^2 = \frac{1}{K}\sum_{k=1}^K \big\|z_k^+ - \Phi(z_k)\widehat{\xi}\big\|^2 + \|\xi - \widehat{\xi}\|_{\widehat{G}}^2,$$

so minimizing it over $\mathcal{C}$ is the same as projecting $\widehat{\xi}$ onto $\mathcal{C}$ in the metric $\|\cdot\|_{\widehat{G}}$. The metric projection onto a closed convex set in a Hilbert space is 1-Lipschitz. Now, using $\bar{\xi} := \Pi_\mathcal{C}(\xi^\star)$, we obtain

$$\|\Pi_\mathcal{C}(\widehat{\xi}) - \xi^\star\|_{\widehat{G}} \le \|\Pi_\mathcal{C}(\widehat{\xi}) - \bar{\xi}\|_{\widehat{G}} + \|\bar{\xi} - \xi^\star\|_{\widehat{G}} \le \|\widehat{\xi} - \xi^\star\|_{\widehat{G}} + \mathrm{dist}_{\widehat{G}}(\xi^\star, \mathcal{C}).$$

If $\xi^\star \in \mathcal{C}$, then $\bar{\xi} = \xi^\star$ and the distance term vanishes. □

Because $\widehat{\xi}_{\mathcal{C}}$ is a projection, the obtuse-angle property $\langle \widehat{\xi} - \widehat{\xi}_{\mathcal{C}},\, \zeta - \widehat{\xi}_{\mathcal{C}} \rangle_{\widehat{G}} \le 0$ holds for every $\zeta \in \mathcal{C}$, which gives the Pythagorean inequality

$$\|\widehat{\xi}_{\mathcal{C}} - \zeta\|_{\widehat{G}}^2 + \|\widehat{\xi} - \widehat{\xi}_{\mathcal{C}}\|_{\widehat{G}}^2 \le \|\widehat{\xi} - \zeta\|_{\widehat{G}}^2 \qquad \forall\, \zeta \in \mathcal{C}.$$

Thus the constrained estimate is never farther than EDMD from any passive reference model, and it is strictly closer unless $\widehat{\xi} \in \mathcal{C}$. This is a deterministic statement and needs no assumption on $\xi^\star$.

The norm in Proposition 5.1 is the data-weighted norm induced by the regression, not the Frobenius norm of the parameter matrices. This statement applies to the pure least-squares problem. Additional penalties, such as those imposed for output mismatch, modify the projection. The term $\operatorname{dist}_{\widehat{G}}(\xi^\star, \mathcal{C})$ is the *structural bias*. Proposition 5.1 thus reduces the statistical question to a deterministic approximation-theoretic one, namely the size of the structural bias. This proposition is consistent with the robustness of constrained fits in the small-sample regime, because the variance contribution is not amplified, whereas the bias is a fixed, data-independent quantity.

Proposition 5.2 (Linear growth of the prediction error)**.** *Let $A_d \in \mathcal{S}_{\mathrm{pH}}(Q)$, let $\hat{z}_k = \Psi(x_k)$ be the lifted true trajectory and $r_k = \hat{z}_{k+1} - A_d \hat{z}_k$ the one-step residual. The model trajectory $z_{k+1} = A_d z_k$ with $z_0 = \hat{z}_0$ satisfies the inequality $\|z_k - \hat{z}_k\|_Q \le \sum_{j<k} \|r_j\|_Q$. If instead only $\|A_d\|_Q = g > 1$ is known, the bound becomes $\sum_{j<k} g^{k-1-j} \|r_j\|_Q$.*

Proof. With $e_k = z_k - \hat{z}_k$ one has $e_{k+1} = A_d e_k - r_k$ and $e_0 = 0$. Therefore, we have the inequality $\|e_{k+1}\|_Q \le \|A_d\|_Q \|e_k\|_Q + \|r_k\|_Q$. □

The bound concerns the lifted error and can be loose, but it is informative when the residuals are small. This explains why unconstrained models with a Q-gain above one can leave the data region: their bound grows geometrically.

6. Choice of the Storage Metric

The lifted pH triple $(J_{\mathrm{K}}, R_{\mathrm{K}}, Q)$ is determined by the data only up to congruence. Therefore, the storage metric Q must be fixed by normalization. The physically meaningful choice is *Hamiltonian consistency*, which is defined as $\frac{1}{2}\Psi^\top Q \Psi = H$, making H_{K} the physical energy. When the Hamiltonian H is known, the admissible Q form the solution set to an SDP feasibility problem.

For the damped pendulum with $H = \frac{1}{2}p^2 + \kappa - \cos\theta$, $\kappa > 1$, and the *dictionary*

$$(1, \cos\theta, \sin\theta, p, p\cos\theta, p\sin\theta),$$

an explicit choice is given by

$$Q = \begin{bmatrix} 2\kappa - 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \oplus \operatorname{diag}(1-q,\, q,\, q), \quad \text{where} \quad 0 < q < 1. \tag{9}$$

This matrix Q is positive definite and satisfies $\frac{1}{2}\Psi^\top Q \Psi = H$ because $\cos^2\theta + \sin^2\theta = 1$. The constant κ is harmless, since only ∇H enters (1). If H is unknown, a pragmatic alternative is to take Q from a Lyapunov certificate of the unconstrained EDMD matrix [28], or to alternate between the convex problem in A_d for fixed Q and an SDP in Q for fixed A_d. Each subproblem is convex, but the alternation carries no global guarantee.

Two choices were tested. (C1) *Energy-anchored*: coordinates $z = (z_E, z_X)$ with $H = \frac{1}{2}|z_E|^2$ and $Q = \operatorname{blkdiag}(I, \alpha I)$, so that $\frac{1}{2} z^\top Q z = H + \frac{\alpha}{2}|z_X|^2$. For the pendulum with $s = \sin(\theta/2)$, $H = \frac{1}{2}p^2 + 2s^2$ is exactly quadratic in (s, p), and $z_E = (p, 2s)$. (C2) *Data-adapted*: Q_L is the solution of $A_c^\top Q_L A_c - Q_L = -I$ with $A_c = \widehat{A} / \max\{1, 1.001\rho(\widehat{A})\}$, rescaled such that $\operatorname{tr} Q_L = N$.

Choice (C1) preserves the physical meaning of the storage, but it carries a large structural bias, because the lifted operator is non-normal and generally not a Q-contraction for such a Q. Choice

(C2) makes $\widehat{A}$ almost a Q_L-contraction, so the bias is small. However, in this case, the physical energy is not represented by the quantity $\frac{1}{2}z^\top Q_L z$ and the guarantee concerns this fitted storage.

Section 7 quantifies this trade-off. The explicit Q in (9) was not benchmarked; the pendulum experiments below use the monomial dictionary in (s, p).

7. Numerical Examples

We compare (a) EDMD (3), (a′) EDMD with the spectrum clipped to the unit disc, (b) restricted EDMD on $\mathcal{S}_{\mathrm{pH}}(Q)$, (c) restricted EDMD with Casimir constraints, and, for the pendulum, (d) restricted EDMD with the data-adapted metric Q_L and (e) a continuous-time pH fit from finite-difference derivatives followed by a Cayley step. We solve the SDPs with CVXPY [29], using Clarabel for $N \leq 40$ and SCS beyond. If the returned matrix has a norm of $\|A_d\|_Q$ slightly above one, we divide it by this gain, except when Casimir constraints are active. In that case the residual is reported. Unless otherwise stated, we report medians over training sets and 30 held-out initial conditions. The error is the relative root mean square error of the original state over 50 steps.

Example 7.1 (Pendulum). We consider the pendulum equation $\ddot{\theta} = -\sin\theta - \gamma\dot{\theta}$, where $\gamma = 0.02$, $\Delta t = 0.1$, with observations $w = (\sin\frac{\theta}{2}, \dot{\theta})$, a dictionary of all monomials of degrees 1 to 5 in w ($N = 20$), and initial conditions sampled uniformly from $\theta \in [-2, 2]$ and $\dot{\theta} \in [-1.4, 1.4]$. *Settings:* S1 (20 trajectories of 10 steps), S2 (100 of 15 steps), S3 (as S2 with noise of standard deviation 0.05 on w), 10 training sets each. The energy gain is $\max_t H(t)/H(0)$ over a free rollout of 1500 steps, computed on the energy coordinates.

Table 1 and Figure 1 show the safety side. The EDMD gains energy: in S2 the median peak is 54 times the initial energy (the worst case is about 10^4). Clipping the spectrum eliminates divergence, but not energy growth (the worst case is 12 to 14 times the initial energy). For the energy-anchored contractions, the storage $\frac{1}{2}z^\top Qz$ cannot increase, therefore, the median energy gain is 1.00. However, the physical energy can still briefly increase in individual runs (the worst case is 1.7 in S1), because the extra coordinates z_X return energy to z_E. The accuracy cost is significant, though: with the physical metric, the 50-step error ranges from 0.12 to 0.31 in noise-free settings, compared to 0.03 to 0.05 for the EDMD. CT-pH achieves an accuracy of $\alpha = 1$ in S1 and S2, consistent with both methods searching the same set up to closure. The data-adapted Q_L is close to EDMD in S2. In S3, EDMD, clipped EDMD, and Q_L all reach the same noise-limited error of 0.224. However, the guarantee concerns the fitted storage, which has a physical energy gain of 1.02 and a worst case value of approximately 8 in S2.

Table 1: Pendulum: median 50-step relative error (Err) and median energy gain (Gain). "div." indicates that the rollout diverged. In S1, EDMD diverged in 89 % of the rollouts.

	S1 (20×10)		S2 (100×15)		S3 (noisy)	
Method	Err	Gain	Err	Gain	Err	Gain
EDMD	0.034	div.	0.048	54	0.224	1.01
EDMD, clipped	0.039	1.12	0.047	1.19	0.224	1.01
CT-pH + Cayley	0.246	1.00	0.314	1.00	0.332	1.00
Q-contr., $\alpha = 1$	0.243	1.00	0.314	1.00	0.586	1.00
Q-contr., $\alpha = 10$	0.117	1.00	0.179	1.00	0.540	1.00
Q-contr., Q_L	0.090	1.00	0.053	1.02	0.224	1.01

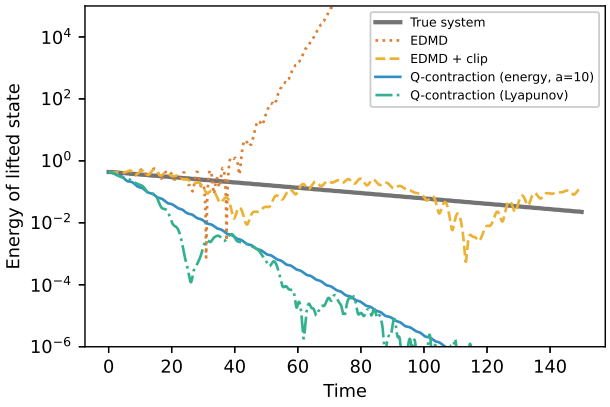


Figure 1: Energy of the lifted energy coordinates along one rollout (S1). The energy-metric contraction exhibits monotonic decay (faster than the true dynamics, indicating structural bias), whereas the Q_L model follows the true behavior more closely, although its physical energy is not monotone.

Figure 2 varies the amplitude of the initial conditions (5 training sets per point). With the additional coordinates at their physical scale, the energy-anchored model achieves 0.6 % error at the smallest amplitude and 56 % at full amplitude. Meanwhile, EDMD and the Q_L model reach approximately 4 % and 4.6 %, respectively. This is consistent with the structural bias $\mathrm{dist}_{\widehat{G}}(\xi^\star, \mathcal{C})$ of Proposition 5.1 growing with the nonlinearity of the data.

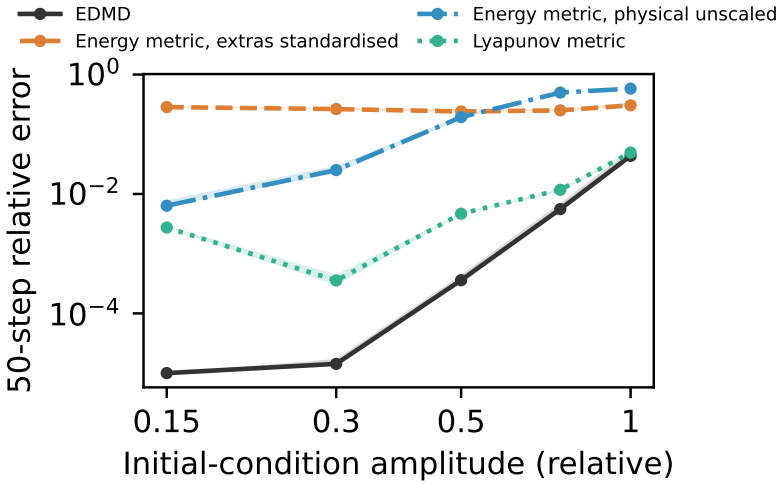


Figure 2: Pendulum, 50-step relative error against the amplitude of the initial conditions (median, interquartile band; errors below 10^{-5} are drawn at 10^{-5}).

Example 7.2 (Damped Rigid Body with a Casimir). For the angular momentum $x \in \mathbb{R}^3$, $H(x) = \frac{1}{2}\sum_i x_i^2/I_i$ with $I = (1, 2, 3)$, and $J(x) = \hat{x}$, the function $C(x) = \frac{1}{2}\|x\|^2$ is a Casimir of J. Along a damped flow, C is conserved only if also $R(x)\nabla C = 0$ (the condition $c \in \ker J \cap \ker R$ in Theorem 3.2(*ii*)). With ordinary damping, the norm $\|x\|^2$ decays and the constraint $c^\top A_d = c^\top$ would be wrong. We therefore use the dissipation $R(x) = \gamma\, \hat{x}^\top \hat{x}$, i.e., $\dot{x} = x \times \nabla H + \gamma\, x \times (x \times \nabla H)$ with $\gamma = 0.5$, for which $\dot{H} = -\gamma\|x \times \nabla H\|^2 \le 0$ and $\dot{C} = 0$. No torque is applied, since with an input $\dot{C} = x^\top B u \neq 0$, and $c^\top B_d = 0$ cannot hold for constant B. The dictionary consists of the monomials of degrees 1 to 3 in x ($N = 19$), which contain H and C exactly, and we use $Q = \mathrm{blkdiag}(I, 10\, I)$ with the energy-anchored coordinates $z_E = (x_i/\sqrt{I_i})_i$.

Settings: R1 (30 trajectories of 10 steps), R2 (100 of 15), R3 (as R2, with noise of standard deviation 0.02 on x); 5 training sets; rollouts of 500 steps.

Table 2: Rigid body: median 50-step error (Err), median energy gain $\max_t H(t)/H(0)$ (Gain), and median Casimir drift $\max_t |C(t) - C(0)|/C(0)$ (Cas.) over 500 steps.

	R1 (30×10)			R2 (100×15)			R3 (noisy)		
Method	Err	Gain	Cas.	Err	Gain	Cas.	Err	Gain	Cas.
EDMD	0.31	$4{\cdot}10^{13}$	$2{\cdot}10^{-5}$	0.22	$8{\cdot}10^{7}$	$8{\cdot}10^{-10}$	0.22	$2{\cdot}10^{5}$	76
EDMD, clipped	0.37	12.7	10^{-11}	0.28	2.5	$2{\cdot}10^{-12}$	0.26	1.17	0.71
Q-contr.	0.35	1.00	1.08	0.32	1.00	1.10	0.37	1.00	1.05
Q-contr. + Casimir	0.34	1.00	$7{\cdot}10^{-9}$	0.33	1.00	$2{\cdot}10^{-8}$	0.37	1.00	$3{\cdot}10^{-8}$

Three observations follow from Table 2. First, the Casimir constraint matters only for noisy data. With exact data, the identity $c^\top Z_Y = c^\top Z_X$ holds, so the least-squares matrix already satisfies $c^\top \widehat{A} = c^\top$ when Z_X has full row rank, and EDMD preserves C (R1, R2). What EDMD does not control is the energy (gains up to 10^{13}; 41 % of the R1 rollouts diverged). With noise (R3), EDMD loses the Casimir completely (median drift 76), and clipping does not restore it.

Second, a Q-contraction without the constraint does not preserve C (drift above 100 %), since nothing in the contraction condition ties C to its initial value.

Third, the Casimir constraint costs no accuracy (0.34 versus 0.35 in R1) and reduces the drift to the order of 10^{-8}. This is solver accuracy (Clarabel), not machine precision, and the contraction property holds with gain one up to the same tolerance. As for the pendulum, all constrained models remain less accurate than EDMD (errors 0.32 to 0.37 versus 0.22 in R2 and R3).

Example 7.3 (Forced Chain with Ports). Two unit masses between walls are coupled by hardening springs with forces $f_j = \delta_j + \delta_j^3$ in the elongations δ_j and damping 0.08; a force input acts on the first mass and is held constant over each step of length 0.1. The dictionary consists of the monomials of degrees 1 to 3 in (x, p) ($N = 34$). The supply is the work $u_k\, F(z_{k+1} - z_k)$, where the row vector F extracts the displacement of mass 1 from z, so that $C_d = F(A_d - I)$ and $D_d = FB_d$, and (8) is linear in (A_d, B_d). Training uses 150 trajectories of 12 steps with random inputs in $[-0.5, 0.5]$, without noise (F1) or with noise 0.02 (F2), 5 training sets. EDMD violates (8) strongly, whereas the smallest eigenvalue of the LMI matrix of the constrained fit is $-7.5 \cdot 10^{-5}$ (F1) and $-7.8 \cdot 10^{-5}$ (F2). This residual comes from the first-order solver, so passivity holds only numerically here. The 100-step error is 0.58 (F1) and 0.52 (F2) for the EDMD and 0.79 for the constrained model. Neither model is accurate, which indicates that this dictionary and data size are too small. The free-decay rollouts of the unforced models did not diverge for any method. We report this example as a check that the port LMI can be imposed, not as evidence of a benefit.

8. Conclusions

For a fixed storage metric, the discrete-time Koopman matrices that admit a lifted pH realization are precisely the Q-contractions, which are an LMI-representable set. Therefore, structure-preserving EDMD is a convex program. Its solution is passive and preserves the imposed linear Casimirs. The constraint cannot increase the estimation error when the population target is pH-compatible. Experiments confirm these guarantees and demonstrate their cost. With physical energy as the storage metric, structural bias can dominate the error. However, a data-adapted metric removes most of the bias but loses the physical meaning of the storage. Open questions include the consistency of the discrete-time estimator as $\Delta t \to 0$, the characterization of discrete-time bilinear pH maps, and the modeling error incurred by identifying a Q-contraction instead of the sampled Koopman matrix, which need not be a Q-contraction for finite Δt.

Acknowledgments

M. Ehrhardt acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 531152215, CRC 1701.

Declarations

Data Availability

Code and data are available from the authors upon request.

Conflict of Interest

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

References

[1] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlin. Sci.*, 25(6):1307–1346, 2015.

[2] M. Korda and I. Mezić. On convergence of extended dynamic mode decomposition to the Koopman operator. *J. Nonlin. Sci.*, 28(2):687–710, 2017.

[3] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proc. Nat. Acad. Sci.*, 17(5):315–318, 1931.

[4] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlin. Dyn.*, 41(1-3):309–325, 2005.

[5] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz. Modern Koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, 2022.

[6] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A Koopman-PINN framework for epidemic models: Parameter inference and forecasting. *Numer. Math. Theor. Meth. Appl.*, 2026.

[7] M. Korda and I. Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.

[8] D. Bruder, X. Fu, R. B. Gillespie, C. D. Remy, and R. Vasudevan. Data-driven control of soft robots using Koopman operator theory. *IEEE Trans. Robot.*, 37(3):948–961, 2021.

[9] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A Koopman operator framework for nonlinear epidemic dynamics: Application to an SIRSD model. *Nonlin. Sci.*, 6:100095, 2026.

[10] A. van der Schaft and D. Jeltsema. Port-Hamiltonian systems theory: An introductory overview. *Found. Trends Syst. Contr.*, 1(2-3):173–378, 2014.

[11] A. van der Schaft. *L_2-Gain and passivity techniques in nonlinear control*. Communications and Control Engineering. Springer International Publishing, 3 edition, 2017.

[12] R. Morandin, J. Nicodemus, and B. Unger. Port-Hamiltonian dynamic mode decomposition. *SIAM J. Sci. Comput.*, 45(4):A1690–A1710, 2023.

[13] K. Cherifi, P. Goyal, and P. Benner. A non-intrusive method to inferring linear port-Hamiltonian realizations using time-domain data. *Electr. Trans. Numer. Anal.*, 56:102–116, 2022.

[14] T. Zwerschke, M. Ehrhardt, M. Günther, and C. Totzeck. Structure-preserving data-driven identification of port-Hamiltonian differential-algebraic systems, 2026. arXiv preprint 2609.07187.

[15] V. M. Preciado. Koopman generator decomposition for port-Hamiltonian systems, 2026. arXiv preprint 2603.17849.

[16] V. Talasila, J. Clemente-Gallardo, and A. J. van der Schaft. Discrete port-Hamiltonian systems. *Syst. Contr. Lett.*, 55(6):478–486, 2006.

[17] P. Kotyczka and L. Lefèvre. Discrete-time port-Hamiltonian systems based on Gauss–Legendre collocation. *IFAC-PapersOnLine*, 51(3):125–130, 2018.

[18] P. Kotyczka and L. Lefèvre. Discrete-time port-Hamiltonian systems: A definition based on symplectic integration. *Syst. Contr. Lett.*, 133:104530, 2019.

[19] V. Mehrmann and R. Morandin. Structure-preserving discretization for port-Hamiltonian descriptor systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, page 6863–6868. IEEE, 2019.

[20] E. Celledoni and E. H. Høiseth. Energy-preserving and passivity-consistent numerical discretization of port-Hamiltonian systems, 2017. arXiv preprint 1706.08621.

[21] L. Lortie, S. Dahdah, and J. R. Forbes. Forward-backward extended DMD with an asymptotic stability constraint, 2024. arXiv preprint 2403.10623.

[22] G. Mamakoukas, I. Abraham, and T. D. Murphey. Learning stable models for prediction and control, 2020. arXiv preprint 2005.04291.

[23] R. Singh, A. Singh, and J. Keshavan. Port-Hamiltonian Koopman operator synthesis for mechanical systems, 2026. arXiv preprint 2609.17249.

[24] J. C. Willems. Dissipative dynamical systems part I: General theory. *Arch. Rat. Mech. Anal.*, 45(5):321–351, 1972.

[25] J. E. Marsden and T. S. Ratiu. *Introduction to mechanics and symmetry: A basic exposition of classical mechanical systems*. Texts in Applied Mathematics. Springer New York, 1999.

[26] O. Gonzalez. Time integration and discrete Hamiltonian systems. *J. Nonlin. Sci.*, 6(5):449–467, 1996.

[27] S. Chen, I. Mezić, and S. Tang. Beyond invariant dictionary: Data-driven koopman spectral recovery with filtered extended dynamic mode decomposition. *arXiv preprint arXiv:2608.02661*, 2026.

[28] J. Zhang and D. Tian. Error amplification in Koopman predictive control: Computable bounds, a conditional Lyapunov certificate, and horizon-consistent estimation. *Mathematics*, 14(21), 2026.

[29] S. Diamond and S. Boyd. CVXPY: A python-embedded modeling language for convex optimization, 2016. arXiv preprint 1603.00943.