LEARNING CONTACT MAPS: UNIVERSALITY AND INVERSION THROUGH FOLDED FRONTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Symplectic map networks combine exact geometric structure with approximation guarantees. We develop the corresponding theory for contact maps, which also represent dissipative and action-dependent dynamics, and connect it to learned inversion. We propose two architectures: LEGENDRENET composes strict lifts of Hénon-like symplectic maps, while CONTACTNET alternates these lifts with prolongations of invertible base maps. We prove that LEGENDRENET is dense in smooth strict contact maps on compact sets. For general contact Hamiltonian flow maps, we establish a modular approximation theorem: any strict family that is dense in C^r, combined with prolongations of any base-map family that is dense in C^{r+1}, gives C^r approximation, with nonsingular factors on the relevant compact sets and convergence of local inverses. This turns contact splitting into a criterion for choosing architectural blocks. Smooth neural-flow bases satisfy the theorem's hypotheses, while triangular bases provide algebraic inverses for recovery. We benchmark the architectures against structure-preserving and generic baselines on damped Hamiltonian systems to quantify finite-capacity approximation. Finally, we use the contact structure of the generalized Hamilton–Jacobi equation to enable recovery of initial smooth profiles from terminal folded fronts.

1 INTRODUCTION

From observed dynamical states, an invertible learned map can support both forward prediction and backward tracing. Structure-preserving architectures encode geometric relations throughout both operations. For conservative systems, SympNets and Hénon-based networks combine exact symplectic structure with approximation guarantees within the symplectic class (Jin et al., 2020; Burby et al., 2021). They restrict the fitted evolution to an expressive family of geometrically admissible maps. We seek the corresponding architecture and approximation theory for contact dynamics, which also include dissipation.

The contact state comprises position x, momentum p, and action u, with $\alpha = du - p \cdot dx$ linking the value and slope of a profile $u = u(x)$. Contact evolution preserves α up to a positive factor (Arnold, 1978; Bravetti, 2017), retaining this compatibility while allowing effects such as friction. Strict contact dynamics recovers the symplectic case, with (x, p) independent of u; general contact dynamics lets the action feed back into their evolution. The architectural question is how to represent this larger class without giving up its geometry or inverse.

Contact variational and splitting integrators construct geometric updates for a specified Hamiltonian, although few generic contact systems admit closed-form splittings (Vermeeren et al., 2019; Bravetti et al., 2020; McLachlan & Quispel, 2002). Geometric Contact Flows (GCF) (Testa et al., 2025) instead learn compositions of tractable contact Hamiltonian subflows, but have no density result for general contact flows. To our knowledge, this is the first neural-architecture density result for general contact Hamiltonian flow maps. Our theorem turns local contact splitting into a modular architecture criterion connecting symplectic and general contact approximation. Appendix I.1 situates this approach among structured dissipative models and generator-learning methods.

The criterion combines *strict lifts of symplectic maps* with *prolongations of invertible base maps*: a C^r-dense strict family and a C^{r+1}-dense base-flow family yield C^r approximation of general contact flows. The extra base derivative is essential. A strict action primitive recovers the derivative spent

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

in forming the pullback, whereas prolongation differentiates the base map to transport momentum; C^{r+1} control also preserves nonsingularity on the relevant compact sets.

Our two architectures instantiate this criterion. LEGENDRENET strictly lifts Hénon-like symplectic updates, while CONTACTNET alternates these lifts with prolonged base maps to allow action-dependent feedback. Both are exactly contact at finite depth on regular domains. Smooth neural-flow bases meet the approximation hypotheses; triangular bases provide algebraic inverses for front recovery.

Hamilton–Jacobi fronts give a concrete use for the jet-state map: their evolution is smooth in (x, u, p) even when the observed (x, u) front folds. A complete ordered front retains its branches, and in one dimension contact compatibility determines the missing momentum across generic cusps. A reconstructed terminal jet and frozen learned inverse then recover the initial profile while preserving the value–slope relation for an exact lift.

Our contributions are:

- **Approximation and architecture criterion.** We prove strict LEGENDRENET density and modular universality for general contact Hamiltonian flow maps, including derivative requirements, nonsingularity, and convergence of local inverses.
- **Inverse problem.** We develop a recovery method for complete ordered folded fronts, supported by momentum-identification and cusp-sensitive stability lemmas, and measure initial-jet recovery and sensitivity to observation noise.
- **Tests of the theory.** We test strict versus general contact dynamics, lifting across symplectic backbones, geometry under sparse training, and inverse recovery.

2 CONTACT DYNAMICS AND THE LEARNING PROBLEM

Contact dynamics. We work on $J^1(\mathbb{R}^n) \simeq \mathbb{R}^{2n+1}$ with states $z = (x, u, p)$, where $x, p \in \mathbb{R}^n$ and $u \in \mathbb{R}$, and with contact form $\alpha = du - p \cdot dx$.

A smooth map Φ is *contact* if $\Phi^*\alpha = e^{\lambda_\Phi}\alpha$ for some scalar function λ_Φ, called the *conformal factor*. It is *strictly contact*, or simply *strict*, if $\lambda_\Phi = 0$. A smooth, invertible contact map is called a *contactomorphism*. Thus, a contact map preserves the contact structure $\ker(\alpha)$, while a strict contact map preserves the contact form α itself. When invertibility is clear from context, we also refer to a contactomorphism simply as a contact map. In particular, a contactomorphism maps a smooth Legendrian submanifold, that is, an n-dimensional submanifold tangent to $\ker(\alpha)$, to another Legendrian submanifold. For a lifted profile, the vanishing of the contact form expresses compatibility between its value and its gradient. Contact maps preserve this relation as position, action, and momentum evolve together, including when the image can no longer be represented by a single-valued profile.

For a smooth contact Hamiltonian $H(x, u, p)$, the state evolves according to the contact Hamilton equations:

$$\dot{x} = \partial_p H, \qquad \dot{p} = -\partial_x H - p\,\partial_u H, \qquad \dot{u} = p \cdot \partial_p H - H. \tag{1}$$

The associated flow map Φ_H^h, obtained after an interval h, is contact wherever the flow is defined (Geiges, 2008; Bravetti, 2017). If H is independent of u, the (x, p) dynamics are symplectic and independent of u. The u equation is then driven by the solution for (x, p), and the resulting map is strict. In contrast, dependence on u can produce friction. For $H = (p^2 + \omega^2 x^2)/2 + \gamma u$ in one dimension with $\gamma > 0$, $\dot{x} = p$ and $\dot{p} = -\omega^2 x - \gamma p$, which is the contact description of the damped harmonic oscillator. The action coordinate thus extends symplectic dynamics from recording an action to allowing action-dependent feedback. Smooth dissipative flows remain invertible on their finite-time flow domains, although backward evolution can amplify perturbations.

Learning the map. Fix a sampling interval $h > 0$. From observed state trajectories we form one-step pairs $\mathcal{D}_h = \{(z_i(t_j), z_i(t_j + h))\}$ and fit a map $\Phi_\theta \approx \Phi_H^h$ directly, without estimating H. Repeated composition gives a predicted state $\hat{z}_N = \Phi_\theta^N(z_0)$. Once θ is fitted from forward data, the same map's inverse gives $\hat{z}_0 = \Phi_\theta^{-N}(z_N)$ from a terminal state z_N generated by the reference dynamics. No separate inverse is trained. This is a reference-state recovery test, while the self-cycle $\Phi_\theta^{-N}(\Phi_\theta^N(z_0))$

tests only the model's own invertibility. We seek architectures that retain this geometric compatibility on their nonsingular domains while learning the evolution and its inverse from forward observations.

3 Invertible Contact Architectures

Strict blocks. The basic LegendreNet atom lifts a Hénon-like symplectic map to the action coordinate. A related lift was used in equation-of-state modeling (Kevrekidis et al., 2024). We denote this atom by $L_{f,\eta,\zeta}$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth trainable potential and $\eta \in \mathbb{R}^n$ and $\zeta \in \mathbb{R}$ are shifts. It sends (x, u, p) to (X, U, P) by

$$L_{f,\eta,\zeta}(x, u, p) = (X, U, P) := \big(p + \eta,\ u - \langle x, p\rangle + f(p) + \zeta,\ -x + \nabla f(p)\big). \tag{2}$$

The (x, p) component is precisely the Hénon-like symplectic map and does not depend on u. Its lift supplies the compatible action update from the same potential, with a free constant shift. The action update contains the pairing $\langle x, p\rangle$ and the potential f, resembling a classical Legendre transformation and motivating the name. Direct substitution gives $L^*_{f,\eta,\zeta}\alpha = \alpha$. The inverse is

$$L^{-1}_{f,\eta,\zeta}(X, U, P) = \big(\nabla f(X - \eta) - P,\ U + \langle \nabla f(X - \eta) - P, X - \eta\rangle - f(X - \eta) - \zeta,\ X - \eta\big). \tag{3}$$

In the 1D strict blocks, zero potential and shifts make an atom a pure Hénon swap, and four such swaps compose to the identity. Groups of four therefore give the strict blocks an identity initialization.

LegendreNet. Given m atoms with parameters $\theta = \{(f_j, \eta_j, \zeta_j)\}_{j=1}^m$, define the corresponding strict architecture network by $\mathcal{L}_\theta = L_{f_m,\eta_m,\zeta_m} \circ \cdots \circ L_{f_1,\eta_1,\zeta_1}$.

The definition allows arbitrary depth m; multiples of four are used only when identity initialization by four-swap groups is desired.

Prolonged base maps. An invertible change of position and action induces a change of slope. Let $B(x, u) = (X(x, u), U(x, u))$ be a smooth invertible map of the base variables. Its first prolongation determines the output momentum P from the generalized chain rule for the transformed slope (Olver, 1995):

$$\mathrm{pr}(B)(x, u, p) = (X, U, P), \qquad P = (U_x + U_u p)(X_x + X_u p)^{-1}. \tag{4}$$

Here p is a row vector, and subscripts denote differentiation with respect to the corresponding variable. On a domain where $X_x + X_u p$ is nonsingular, $\mathrm{pr}(B)^*\alpha = (U_u - PX_u)\alpha$. Thus an invertible base map alone does not guarantee that its prolongation is defined for every momentum. For the 1D inverse experiments, we choose triangular base blocks to make the base inverse algebraic. A representative block first sets $U = a(x)u + b(x)$ and then $X = c(U)x + d(U)$, where a, b, c, d are trainable functions and the scales a, c are positive. Their positivity guarantees algebraic invertibility of the base substeps; prolongation regularity still requires the denominator condition above. After inverting the base relations, the old momentum follows from $p = (PX_x - U_x)/(U_u - PX_u)$ at the recovered (x, u). Higher-dimensional blocks use u-conditioned spatial maps and x-conditioned action shears. Because Equation (4) differentiates B, approximation of a prolonged map consumes one derivative of the base approximation. Full block formulas are deferred to the appendix.

These base transformations supply action-dependent feedback beyond the strict class. For example, the base map $X = x$, $U = e^{-\gamma h}u$ prolongs to $P = e^{-\gamma h}p$, the flow over an interval h of $H = \gamma u$. Composed with strict factors, it models damping together with conservative motion.

Full architecture and inverse. For either ContactNet variant, write L_j for a LegendreNet block, B_j for an invertible base map, and $C_j = \mathrm{pr}(B_j)$ for its prolongation. One stage applies L_j, then C_j, and, when enabled, a vertical action block V_j (Figure 1).

ContactNet. Given m stages with trainable parameters θ, define the contact architecture by

$$\Phi_\theta = (V_m \circ C_m \circ L_m) \circ \cdots \circ (V_1 \circ C_1 \circ L_1), \qquad V_j = \mathrm{id} \text{ when omitted.} \tag{5}$$

The optional vertical block in triangular ContactNet has $X = x$, $U = u + \rho F(x, u)$, and $P = U_x + U_u p$, where F is a trainable scalar field and ρ is its scale. Its scalar inverse solves $u = U - \rho F(X, u)$ by fixed-point iteration, then recovers p algebraically. Convergence follows

when $\sup |\rho F_u| < 1$ on the relevant domain. The strict atoms and triangular base blocks invert algebraically, while the vertical block uses this iterative step. Reversing the stage order gives the inverse of the full fitted map Φ_θ, with no separately trained decoder.

The two constructions therefore provide composable contact maps with inverses obtained from their fitted factors. Every finite composition preserves the contact distribution wherever its prolongation denominators and action Jacobians are nonsingular. In practice, hinge penalties on the absolute prolongation denominators and, for residual vertical blocks, their action Jacobians encourage margins from singularity on sampled states.

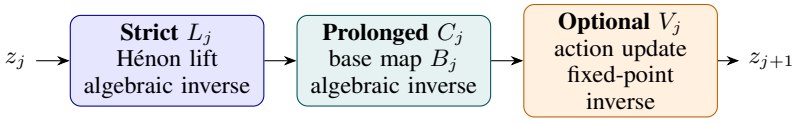

Figure 1: A triangular-base CONTACTNET stage. The strict block updates the symplectic variables and lifts the action. The prolonged block introduces action-dependent feedback. The vertical block is optional. The inverse evaluates the blocks from right to left.

4 APPROXIMATION THEORY

The approximation question is whether enforcing contact structure leaves the architecture expressive within the intended contact class. We establish this for strict maps and contact Hamiltonian flow maps using Turaev's Hénon approximation (Turaev, 2003) and the strict/prolonged splitting of contact flows (Kevrekidis, 2026). Our dynamical targets are finite-time flows on $J^1(\mathbb{R}^n)$, in the identity component of the contact maps. The results concern compact-flow domains in this jet space; other components and general contact manifolds are outside their scope.

Strict universality. The first result concerns the LEGENDRENET class of Section 3, with depth and potential capacity allowed to grow with the requested accuracy.

Theorem 4.1 (Strict LEGENDRENET universality)**.** *Let $\Phi : J^1(\mathbb{R}^n) \to J^1(\mathbb{R}^n)$ be a smooth strict contactomorphism, and let $K \subset J^1(\mathbb{R}^n)$ be compact. For every integer $r \geq 0$ and every $\varepsilon > 0$, there is a* LEGENDRENET *$\mathcal{L}_\theta$ such that $\|\mathcal{L}_\theta - \Phi\|_{C^r(K)} < \varepsilon$.*

The proof is given in Appendix B. We briefly outline the argument. A strict contactomorphism preserves the Reeb vector field ∂_u, so it has the form $\Phi(x, u, p) = (\phi(x, p), u + A_\phi(x, p) + c)$, where ϕ is symplectic and $dA_\phi = \phi^*(p \cdot dx) - p \cdot dx$ (Geiges, 2008; Kevrekidis, 2026). The atoms in Equation (2) provide the corresponding strict lifts, with the shift ζ supplying the constant c. Turaev's Hénon approximation provides the required approximations of ϕ in the (x, p) variables. Lifting these approximations then gives the desired LEGENDRENET. For $r \geq 1$, approximation of ϕ in C^r controls the action primitive, normalized at a common point, and hence the strict lift in C^r: forming the pullback differentiates the map once, while taking its primitive recovers that derivative. For the C^0 conclusion, we use C^1 approximation of ϕ. Thus the compatible action update inherits expressivity from the symplectic approximation, with its additive constant supplied by the lift.

Modular contact universality. The local universal splitting result of Kevrekidis (2026) expresses a contact Hamiltonian flow map locally as a limit of compositions of strict flows and first prolongations of base flows. Replacing those exact factors with trainable ones requires approximation in the right topology.

We say that a collection of maps $\mathcal{A}$ is an *r-approximator* for a class of maps $\mathcal{C}$ if, for every $\Phi \in \mathcal{C}$, every compact set K in the domain of Φ, and every $\varepsilon > 0$, there is a map $\Psi \in \mathcal{A}$ such that $\|\Psi - \Phi\|_{C^r(K)} < \varepsilon$. If this holds for every finite $r \geq 0$, we simply call $\mathcal{A}$ an *approximator* for $\mathcal{C}$.

Theorem 4.2 (Modular contact approximation)**.** *Let $r \geq 0$ be an integer. Let $\mathcal{A}$ be an r-approximator for smooth strict contactomorphisms, and let $\mathcal{B}$ be an $(r+1)$-approximator for smooth base-flow maps on $\mathbb{R}^n \times \mathbb{R}$. Suppose that the map Φ_H^h generated over an interval h by a smooth contact Hamiltonian is defined on an open neighborhood of a compact set K. Assume also that the subflows used in the local splitting construction are defined on neighborhoods of all compact sets encountered in the composition.*

Then, for every $\varepsilon > 0$, there is a finite alternating composition of maps from $\mathcal{A}$ and prolongations of maps from $\mathcal{B}$ that approximates Φ_H^h within ε in $C^r(K)$. The prolonged factors are well-defined and nonsingular on the compact sets encountered in the approximation.

The proof in Appendix C fixes a finite splitting within half the tolerance and approximates each strict and base-flow factor on compact neighborhoods of the intermediate states. The extra base derivative is what controls the Jacobians in Equation (4), since value-level approximation of B does not control the slopes that transport momentum. The theorem consequently gives a reusable design criterion: different strict and base-map families inherit contact approximation when they meet the stated derivative requirements.

Inverse approximation. For $r \geq 1$, C^r convergence of diffeomorphisms on a compact neighborhood of K implies C^r convergence of their local inverses on every compact subset of $\Phi_H^h(\operatorname{int} K)$, provided the target Jacobian is nonsingular there. The inverse function theorem and differentiation of $F^{-1} \circ F = \mathrm{id}$ give this conclusion. It is qualitative. The constants depend on the target inverse Jacobian, so approximation of the inverse does not bound the amplification of a terminal perturbation.

Corollary 4.3 (Universal flow-based and algebraically invertible triangular families)**.** *Fix $r \geq 0$ and the compact-flow domains of Theorem 4.2.* ***(a) Universal*** CONTACTNET**.** LEGENDRENET *strict blocks and prolongations of any $(r+1)$-approximating base family are dense in smooth contact Hamiltonian flow maps in C^r.* ***(b) Closed-form-invertible instantiation.*** *The implemented triangular base family supplies blockwise algebraic inverses on regular domains. Its* CONTACTNET *is C^r-dense if its variable-depth base compositions are C^{r+1}-dense in the required base-flow maps. Optional vertical blocks may be omitted in both statements.*

Part (a) establishes a universal contact architecture: exact flows of smooth neural vector fields with tanh activation supply the required base family (Lemma C.3). Part (b) provides a task-oriented alternative with algebraic base inverses. For the triangular family, the stated C^{r+1} base-density condition is not established here; value-level C^0 or L^p density would not suffice. Thus the base blocks can be chosen according to the approximation and inversion guarantees required by the task.

5 FINITE-CAPACITY BEHAVIOR OF CONTACT ARCHITECTURES

The theory has two immediate architectural consequences: strict blocks alone cannot represent targets with a nontrivial contact factor, while strict lifting extends to symplectic families with computable action primitives. We test these consequences at finite capacity, then examine how factorization shapes prediction and recovery. One-step fit, multistep rollout, reference-state recovery, and geometric fidelity measure distinct properties. We compare CONTACTNET and LEGENDRENET with GCF, an invertible contact architecture built from Hamiltonian subflows (Testa et al., 2025); RealNVP, which is invertible without imposing contactness (Dinh et al., 2017); and an unconstrained MLP, which has no architectural inverse. The oscillator and action-transport comparisons also include neural-ODE CONTACTNET (CONTACTNET-ODE), which prolongs a learned base flow on (x, u); the appendix specifies its numerical realization.

Choosing the contact class. On the damped oscillator $H = (p^2 + x^2)/2 + 0.15u$, LEGENDRENET has mean contact defect 2.60×10^{-16} yet final-time state MSE 7.12×10^{-2} (Table 7). Its strict constraint preserves the form exactly but cannot reproduce the target's nontrivial contact factor. The prolonged blocks extend the admissible dynamics beyond the strict class while retaining contact compatibility.

Strict lifts beyond LEGENDRENET**.** The strict side of the modular criterion is not specific to Hénon-like or Legendre blocks. If a symplectic architecture has a computable action primitive A_ϕ satisfying $dA_\phi = \phi^*(p \cdot dx) - p \cdot dx$, then its phase map can be extended to a strict contact map by updating the action as $U = u + A_\phi + c$. Thus strict lifting applies to **any symplectic architecture for which this action primitive is computable**. In Appendix F.1, we benchmark LEGENDRENET against the strict lift of G-SympNet (Jin et al., 2020), together with learned action-head controls.

With 32 free-propagation training jets, the exact lift lowers action MSE from 6.05×10^{-4} to 3.63×10^{-5} on the Hénon backbone and from 5.77×10^{-4} to 2.22×10^{-7} on G-SympNet, while keeping the strict-contact defect at numerical precision (Table 6). Across both backbones, the lift

supplies a compatible action update from the learned phase map. Their different errors reflect the finite parameterization and its fitted phase dynamics.

Benchmark suite. The controlled systems range from free propagation,

$$H_{\text{free}}(x,u,p) = \tfrac{1}{2}p^2, \qquad \Phi^h_{\text{free}}(x,u,p) = \big(x + hp,\ u + \tfrac{h}{2}p^2,\ p\big), \tag{6}$$

through conservative and damped oscillators, polynomial action dependence across depths, and action-dependent transport, to the folded fronts of Sections 6 and 7. Table 1 records the role of each study; protocols and seed-level results are in the appendices.

5.1 Inverse recovery and contact factorization

Compatible inverse recovery. In the quadratic-front inverse problem later developed in Sections 6 and 7.2, RealNVP and ContactNet recover with backward-trajectory MSE 2.56×10^{-4} and 2.75×10^{-4}, while their forward Legendrian residuals are 5.88×10^{-4} and 1.39×10^{-5} (Tables 16 and 18). The recovered initial jets also differ geometrically: ContactNet's residual is about 39 times smaller at similar state error (Section 7.2). These residuals test whether recovered action and momentum obey their derivative relation, which contact maps preserve and generic invertibility does not impose.

Factorization within the contact class. GCF preserves the contact identity to numerical precision, but its quadratic recovery MSE is 5.99×10^{-3}, about 22 times ContactNet's. Its forward fit is also worse. The recovery gap therefore measures the combined error of the fitted dynamics and their architectural inverse rather than inverse conditioning alone. The two contact parameterizations learn different approximations to the reference flow, and their architectural inverses reverse those fitted maps; a small self-cycle instead measures inversion of the map itself.

5.2 Factorization, optimization, and extrapolation

The GCF contactomorphism parameterization composes subflows generated by $M(p)p^2/2$, $V(x)$, and $F(x)u$. Free propagation in Equation (6) is exactly aligned with its kinetic factor. For the oscillator family

$$H_\gamma(x,u,p) = \tfrac{1}{2}(p^2 + x^2) + \gamma u, \qquad \gamma \in \{0, 0.15\}, \tag{7}$$

GCF has the lowest mean final-time state error under the common 80-epoch Adam/40-step L-BFGS schedule: 3.62×10^{-3} and 2.27×10^{-4}, respectively. Together with free propagation, these results show the benefit of a contact factorization aligned with the target.

The controlled four-power/depth sweep changes the winner with both action power and depth (Appendix F.3). An action-transport test favors ContactNet for position response to changing initial action, although the MLP has lower full-state error and the training costs differ (Appendix F.3.1 and Table 5). ContactNet-ODE retains sampled contact precision, with lower mean transport error and higher oscillator error than triangular ContactNet. These results support different finite-depth factorization biases, without a monotone rule linking action power to trained accuracy.

Doubling the oscillator initial-state range produces large errors for every model class and nonfinite rollouts for some (Table 8). The experiments thus identify target- and factorization-specific inductive biases at finite capacity: contact compatibility persists on regular domains, while the fitted dynamics determine prediction and extrapolation accuracy.

6 Hamilton–Jacobi Fronts and the Inverse Problem

Smooth dynamics above a folded front. Front evolution makes the compatibility of value and slope directly useful. The generalized Hamilton–Jacobi equation

$$\partial_t u(x,t) + H\big(x, u(x,t), \nabla_x u(x,t)\big) = 0 \tag{8}$$

connects contact dynamics to front evolution (Guenther et al., 1995; Bravetti, 2017). An initial profile u_0 defines the Legendrian jet $\iota_0(s) = (s, u_0(s), \nabla u_0(s))$. Its characteristic evolution $\iota_t = \Phi^t_H \circ \iota_0$ has a smooth Legendrian image $L_t = \iota_t(\mathbb{R}^n)$ wherever the flow is defined. Where the spatial projection is locally invertible, this image is the one-jet of a classical solution of Equation (8).

Writing $\iota_t = (X_t, U_t, P_t)$, a caustic occurs where $\det D_s X_t = 0$: the projected front (X_t, U_t) can fold over x while the lifted front remains smooth (Arnol'd, 1990) (Figure 2). Our inverse problem uses the complete ordered front, including every branch. A single-valued solution selected by a viscosity or entropy condition, when applicable, can discard branches needed for this recovery (Crandall & Lions, 1983; Evans, 2010).

Relation to PDE learning. DeepONet and Fourier neural operators learn maps between function spaces (Lu et al., 2021; Li et al., 2021). Here the map acts on each ray's jet, with recovery from a complete ordered front as in a ray-resolved wavefront description.

Lifting a one-dimensional terminal front. In $J^1(\mathbb{R})$, the ordered projected curve determines the missing momentum needed by the contact inverse, under a genericity condition.

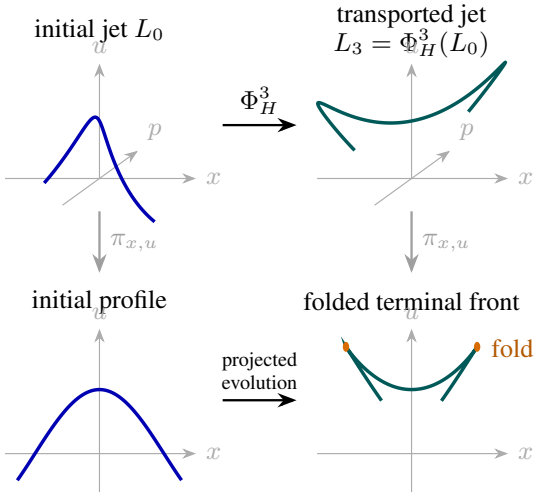

Figure 2: Analytic illustration with $H = p^2/2$ and $u_0(s) = \cos s$: the jet (top) stays smooth under Φ_H^3 while its (x, u) projection (bottom) folds where $X_s = 1 - 3\cos s$ vanishes.

Proposition 6.1 (Lift across generic cusps). *Let $s \mapsto (x(s), u(s), p(s))$ be a C^2 embedded Legendrian curve, and suppose every zero s_0 of x_s satisfies $x_{ss}(s_0) \neq 0$. Then its ordered (x, u) front uniquely determines p by*

$$p(s) = \frac{u_s(s)}{x_s(s)} \quad \text{if } x_s(s) \neq 0, \qquad p(s_0) = \frac{u_{ss}(s_0)}{x_{ss}(s_0)} \quad \text{if } x_s(s_0) = 0. \tag{9}$$

At each cusp s_0, the ratio $u_s(s)/x_s(s)$ has the same limit from either side as $s \to s_0$, and that limit equals $u_{ss}(s_0)/x_{ss}(s_0)$.

The formulas follow from $u_s = p x_s$ and its derivative at a cusp (proof in Appendix H). This establishes identifiability from the complete front. For sampled observations, the pointwise stability bound in Proposition H.1 scales inversely with $|x_s|$, so we estimate the ratio away from small denominators and interpolate momentum across the masked interval, at the price of an interpolation-bias term.

Inverse task. From the reference terminal front at $T = Nh$, we estimate momentum using Equation (9), form $\widehat{\iota}_T = (X_T, U_T, \widehat{P}_T)$, and apply the frozen inverse: $\widehat{\iota}_0 = \Phi_\theta^{-N}(\widehat{\iota}_T)$. We compare this jet with ι_0; its action coordinate recovers the initial profile. No inverse model or initial state is optimized. Here H is unknown to the learner, which is trained on forward full-jet pairs, and projected fronts enter only at inverse evaluation. With H known, characteristics could instead be integrated backward. A contact inverse maps an exactly Legendrian terminal jet to a Legendrian jet, so the recovered slope is the profile derivative wherever the recovered spatial projection is regular. A generic invertible map imposes no such relation among its three output coordinates. In two dimensions, $D_s U = P\, D_s X$ determines momentum where $D_s X$ is invertible, but Proposition 6.1 does not establish continuation across rank-deficient caustics. Our 2D inverse experiment therefore supplies terminal momentum.

Three errors enter: the sampled front gives an approximate terminal momentum, the learned map differs from the reference evolution, and repeated inverse steps can amplify either. We therefore report momentum-lift error, backward-trajectory state error, and the recovered-jet Legendrian residual separately. For an exact contact inverse F, $(F \circ \iota)^* \alpha = \iota^*(F^* \alpha)$, so the last residual vanishes exactly when the lifted terminal curve is Legendrian.

7 Front Learning and Recovery in One and Two Dimensions

We first test how architectural compatibility and fitted state accuracy behave with few training fronts, then reuse separately trained forward maps for inverse recovery. Appendices E and G give the protocols and extended comparisons.

7.1 GEOMETRY WITH FEW TRAINING FRONTS

For $H = p^2/2 + x^2/2 + u^2/4$, we train approximately 2.7k-parameter MLP, RealNVP, and CONTACTNET models on nested sets of 2, 4, 8, or 40 fronts, with three model seeds. Eight validation fronts select checkpoints; 15 disjoint test fronts are all folded at $t = 1.2$. Figure 3 compares terminal state error and sampled Legendrian geometry, including sensitivity to optimization budget.

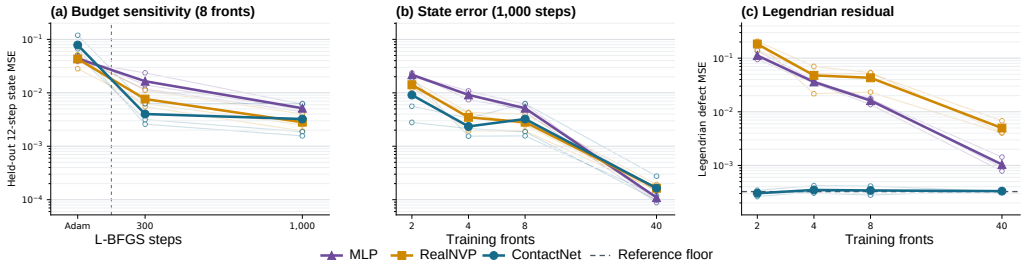

Figure 3: Quadratic-action fronts, three seeds. (a) 12-step terminal state MSE at eight training fronts after 600 Adam updates and after 300 or 1,000 L-BFGS steps. (b–c) State and sampled Legendrian MSE across front counts at 1,000 steps. Thin lines join paired seeds; thick lines show means. The dashed line is the reference finite-difference floor, 3.25×10^{-4}. Segments connect measured settings.

As the contact construction guarantees, up to discretization, panel (c) shows preserved Legendrian geometry: at two fronts CONTACTNET gives 3.03×10^{-4} versus RealNVP's 1.83×10^{-1}. Panels (a–b) test the empirical question of state accuracy: CONTACTNET has the lowest three-seed mean at two and four fronts, RealNVP at eight, and the MLP at forty. Paired rankings are mixed. The 1,000-step comparison includes the MLP, RealNVP, and CONTACTNET; GCF was evaluated only under the 300-step recipe reported in Appendix G.1. The geometric constraint remains active at every data level; training determines which compatible evolution is fitted.

At eight fronts CONTACTNET beats RealNVP in two of three paired seeds but has the worse mean. Data efficiency also depends on how thoroughly each model is trained. The plotted floor is the reference fronts' own finite-difference residual, which depends on the profile set. The seeded study's fronts give 1.38×10^{-5} under the same convention, so residual scales are comparable within a study, not across studies (Appendix E).

7.2 RECOVERY AND GEOMETRIC CONSISTENCY IN ONE DIMENSION

For the quadratic-action system, a separate higher-budget CONTACTNET trained on forward pairs recovers initial jets from complete ordered terminal (x, u) fronts. We estimate momentum by finite differences with interpolation near cusps (Appendix H), then apply 12 frozen inverse steps. Of 15 held-out fronts, 12 are folded at terminal time. Across these 12, median momentum-lift MSE is 5.70×10^{-5} and median backward-trajectory state MSE is 6.29×10^{-5}. These medians are taken across fronts for a single trained model. Figure 4 (left, center) shows one recovery; Appendix G.4 reports all cases. In the three-seed comparison, the recovered initial-jet Legendrian residual after 12 inverse steps is 2.35×10^{-5} for CONTACTNET versus 9.25×10^{-4} for RealNVP (Table 18). The reference jets themselves have residual 1.79×10^{-5} under the same convention, so CONTACTNET's recovered jets sit at the discretization scale and RealNVP's do not, despite comparable state error (Appendix G.2).

Sensitivity to observed front noise. We add independent Gaussian noise to each observed terminal x and u coordinate, three paired draws for each of the 12 folded fronts, with the checkpoint frozen. A smoothed variant first applies a fixed nine-ray, degree-three Savitzky–Golay filter to both coordinates. At noise standard deviation 10^{-3} the median recovery MSE is 3.54×10^{-3} raw and 1.42×10^{-3} smoothed, and at 10^{-2} it is 2.79×10^{-1} and 5.92×10^{-2}, far above the clean baseline (Table 20). Smoothing trades derivative noise for bias, raising clean recovery error from 6.29×10^{-5} to 6.94×10^{-5}. These results expose the sensitivity of the observation-to-jet step, whose errors can then be amplified by the learned inverse.

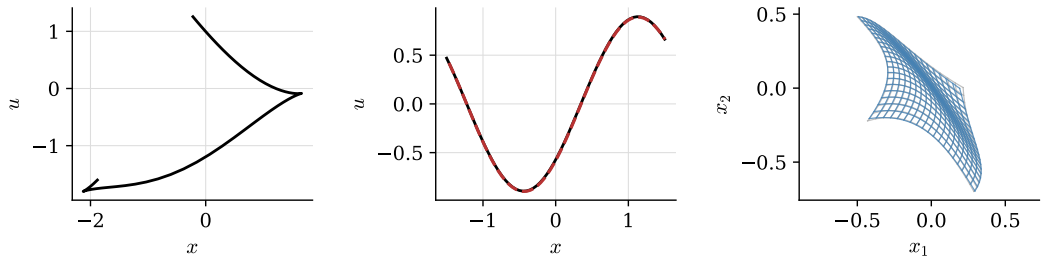

Figure 4: Front recovery and transport with CONTACTNET. Left: held-out 1D front 5, the earliest-folding test front, terminal (x, u) curve. Center: its recovered initial profile (red dashed) over the reference (black). Right: held-out 2D profile 5, the middle test index, predicted terminal ray mesh (blue) over the reference (gray); the 2D full-jet inverse is in Figure 8.

7.3 TRANSPORT AND FULL-JET INVERSION IN TWO DIMENSIONS

The same contact construction acts on Legendrian surfaces. With $H = \|p\|^2/2 + u^2/4$, we train on 30 profiles and evaluate on 11 held-out profiles. Full-state rollout MSE is 6.49×10^{-5} for CONTACTNET versus 9.81×10^{-3} for the MLP, with a smaller Legendrian residual. These are profile averages from one trained model per architecture; checkpoint selection used the same held-out profiles, but the last checkpoints give the same $151\times$ ratio (Appendix G.5). Figure 4 (right) shows a terminal caustic. For one held-out case, 12 frozen inverse steps recover the initial (x, u) front with MSE 4.14×10^{-5} from the supplied reference terminal *full jet*. This isolates the learned inverse from projected-to-jet reconstruction at a caustic, a problem Proposition 6.1 does not cover in two dimensions.

8 LIMITATIONS AND CONCLUSION

Scope. The approximation results are local on compact sets and allow capacity to grow. Neural-flow bases give a proved universal family, realized numerically in the forward comparison; triangular bases serve the inverse experiments with algebraic base inverses. Finite networks preserve contact geometry on regular domains; the sampled penalty margins are not global regularity certificates, and the optional vertical inverse additionally needs a contraction condition. These properties supply neither a prediction-error bound nor a uniform inverse-conditioning bound (Appendix G.3).

The sparse-front experiment measures one Hamiltonian and one nested sequence of training profiles with three model seeds, rather than a general sample complexity law. The seeded five-model front comparison uses reported held-out profiles for checkpoint selection and unequal budgets, although its forward rankings are unchanged at the final checkpoints (Table 17). The 1D recovery assumes complete ordered observations. Its local stability bound deteriorates near cusps; the noise sweep tests only one grid, threshold, smoothing rule, and trained model. Pointwise differentiation of the observed front contributes to this sensitivity, and inverse evolution can amplify the resulting error. A weak-form lift that pairs the front with smooth test functions, in the spirit of weak-form model identification (Messenger & Bortz, 2021), avoids that differentiation and is left to future work. The 2D recovery test covers one case from a supplied terminal full jet.

Conclusion. We developed a modular architecture for learning contact maps, extending structure-preserving map learning from symplectic phase dynamics to the full position–action–momentum state. Strict lifting supplies the compatible action update; composition with prolonged base maps permits action-dependent feedback. The approximation results establish a modular criterion for approximating contact Hamiltonian flow maps on compact-flow domains, with the base family determining the available approximation and inversion guarantees. At finite capacity, exact action lifting improves action accuracy across two symplectic backbones, while the contact architectures keep sampled Legendrian residuals near their discretization floors. Prediction accuracy, however, depends on the target, factorization, and optimization budget. For complete ordered fronts, the same forward-trained map can recover initial profiles through folded projections while retaining value–slope compatibility for exact terminal lifts. Together, the density results and the two architectures connect contact approximation, prediction, and inversion within a single map-learning construction.

Reproducibility Statement

Mathematical statements and accompanying proofs and assumptions appear in Appendices B, C and H. Neural architecture formulas, training protocols, metrics, and extended results appear in Appendices D to G. All data sets used in this work are synthetic. The manuscript contains sufficient information for an independent reproduction of the results. Code associated with this manuscript will become publicly available on GitHub, conditionally upon acceptance.

AI Use Statement

Generative AI tools (OpenAI Codex using GPT-5 and GPT-6 models and Anthropic Claude using Fable 5 and Opus 5 models) were used to draft and edit portions of the manuscript, assist with drafting and checking proofs, create and edit diagnostic and presentation code, search for relevant literature, and perform consistency checks. They were not used to originate the LegendreNet or ContactNet architectures or the core mathematical framework, which predated their use. The ideas, claims, and framing are the authors'. All AI-assisted text, mathematical arguments, citations, and code were reviewed by the authors. The authors take responsibility for the final content of this work, including all text, claims, code, and artifacts produced with the aid of generative AI.

References

V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer New York, 1st edition, 1978. doi: 10.1007/978-1-4757-1693-1.

Vladimir Igorevich Arnol'd. *Singularities of Caustics and Wave Fronts*, volume 62 of *Mathematics and Its Applications*. Kluwer Academic Publishers, 1990. doi: 10.1007/978-94-011-3330-2.

Tom Bertalan, Felix Dietrich, Igor Mezić, and Ioannis G Kevrekidis. On learning hamiltonian systems from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12):121107, 2019. doi: 10.1063/1.5128231.

Shubham Bhardwaj and Chandrajit Bajaj. PHAST: Port-hamiltonian architecture for structured temporal dynamics forecasting. *arXiv preprint arXiv:2602.17998*, 2026. doi: 10.48550/arXiv.2602.17998. URL https://arxiv.org/abs/2602.17998.

Alessandro Bravetti. Contact Hamiltonian dynamics: The concept and its use. *Entropy*, 19(10):535, 2017. doi: 10.3390/e19100535.

Alessandro Bravetti, Marcello Seri, Mats Vermeeren, and Federico Zadra. Numerical integration in celestial mechanics: a case for contact geometry. *Celestial Mechanics and Dynamical Astronomy*, 132(1):7, 2020. doi: 10.1007/s10569-019-9946-9.

J. W. Burby, Q. Tang, and R. Maulik. Fast neural Poincare maps for toroidal magnetic fields. *Plasma Physics and Controlled Fusion*, 63(2):024001, 2021. doi: 10.1088/1361-6587/abcbaa.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31, pp. 6572–6583, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html.

Michael G Crandall and Pierre-Louis Lions. Viscosity solutions of Hamilton–Jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–42, 1983. doi: 10.1090/S0002-9947-1983-0690039-8.

Manuel de León and Manuel Lainz Valcázar. Contact hamiltonian systems. *Journal of Mathematical Physics*, 60(10):102902, 2019. doi: 10.1063/1.5096475.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=HkpbnH9lx.

Valentin Duruisseaux, Joshua W Burby, and Qi Tang. Approximation of nearly-periodic symplectic maps via structure-preserving neural networks. *Scientific reports*, 13(1):8351, 2023. doi: 10.1038/s41598-023-34862-w.

Lawrence C Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2nd edition, 2010. doi: 10.1090/gsm/019.

Hansjörg Geiges. *An introduction to contact topology*, volume 109 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2008. doi: 10.1017/CBO9780511611438.

Jiale Gong, Pengzhan Jin, Dongyang Kuang, Lu Li, and Yifa Tang. CSympNet-ID: Conformal-symplectic map learning for linearly damped hamiltonian systems, 2026. URL https://arxiv.org/abs/2607.03339. arXiv:2607.03339.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in neural information processing systems*, volume 32, pp. 15353–15363, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/26cd8ecadce0d4efd6cc8a8725cbd1f8-Abstract.html.

Miroslav Grmela and Hans Christian Öttinger. Dynamics and thermodynamics of complex fluids. I. development of a general formalism. *Physical Review E*, 56(6):6620–6632, 1997. doi: 10.1103/PhysRevE.56.6620.

R. B. Guenther, C. M. Guenther, and J. A. Gottsch. *The Herglotz lectures on contact transformations and Hamiltonian systems*, volume 1 of *Lecture Notes in Nonlinear Analysis*. Juliusz Schauder Center for Nonlinear Studies. Nicholas Copernicus University, 1995. URL https://wydawnictwo.umk.pl/produkt/the-herglotz-lectures-on-contact-transformations-and-hamiltonian-systems.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2): 251–257, 1991. doi: 10.1016/0893-6080(91)90009-T.

Isao Ishikawa, Takeshi Teshima, Koichi Tojo, Kenta Oono, Masahiro Ikeda, and Masashi Sugiyama. Universal approximation property of invertible neural networks. *Journal of Machine Learning Research*, 24(287):1–68, 2023. URL https://jmlr.org/papers/v24/22-0384.html.

Pengzhan Jin, Zhen Zhang, Aiqing Zhu, Yifa Tang, and George Em Karniadakis. Sympnets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. *Neural Networks*, 132:166–179, 2020. doi: 10.1016/j.neunet.2020.08.017.

George A Kevrekidis. Local universal splitting integrators for contact hamiltonian systems, 2026. URL https://arxiv.org/abs/2605.09103. arXiv:2605.09103.

George A Kevrekidis, Daniel A Serino, M. Alexander R. Kaltenborn, J. Tinka Gammel, Joshua W. Burby, and Marc L. Klasky. Neural network representations of multiphase equations of state. *Scientific Reports*, 14(1):30288, 2024. doi: 10.1038/s41598-024-81445-4.

Patrick Kidger. *On neural differential equations*. PhD thesis, University of Oxford, 2021. URL https://arxiv.org/abs/2202.02435.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.08895.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Junlong Lyu, Zhitang Chen, Chang Feng, Wenjing Cun, Shengyu Zhu, Yanhui Geng, Zhijie Xu, and Yongwei Chen. Universality of parametric coupling flows over parametric diffeomorphisms. *arXiv preprint arXiv:2202.02906*, 2022. doi: 10.48550/arXiv.2202.02906. URL https://arxiv.org/abs/2202.02906.

Robert I McLachlan and G Reinout W Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002. doi: 10.1017/S0962492902000053.

Daniel A Messenger and David M Bortz. Weak SINDy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021. doi: 10.1137/20M1343166.

Peter J Olver. *Equivalence, invariants and symmetry*. Cambridge University Press, 1995. doi: 10.1017/CBO9780511609565.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021. URL https://jmlr.org/papers/v22/19-1028.html.

Benjamin Peherstorfer. Model reduction for transport-dominated problems via online adaptive bases and adaptive sampling. *SIAM Journal on Scientific Computing*, 42(5):A2803–A2836, 2020. doi: 10.1137/19M1257275.

Donsub Rim, Scott Moe, and Randall J LeVeque. Transport reversal for model reduction of hyperbolic partial differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 6(1):118–150, 2018. doi: 10.1137/17M1113679.

Andrew Sosanya and Sam Greydanus. Dissipative hamiltonian neural networks: Learning dissipative and conservative dynamics separately. *arXiv preprint arXiv:2201.10085*, 2022. doi: 10.48550/arXiv.2201.10085. URL https://arxiv.org/abs/2201.10085.

Takeshi Teshima, Isao Ishikawa, Koichi Tojo, Kenta Oono, Masahiro Ikeda, and Masashi Sugiyama. Coupling-based invertible neural networks are universal diffeomorphism approximators. In *Advances in Neural Information Processing Systems*, volume 33, pp. 3362–3373, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2290a7385ed77cc5592dc2153229f082-Abstract.html.

Andrea Testa, Søren Hauberg, Tamim Asfour, and Leonel Rozo. Geometric contact flows: Contactomorphisms for dynamics and control. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 59259–59284, 2025. URL https://proceedings.mlr.press/v267/testa25a.html.

Ch. Tsitouras. Runge–kutta pairs of order 5(4) satisfying only the first column simplifying assumption. *Computers & Mathematics with Applications*, 62(2):770–775, 2011. doi: 10.1016/j.camwa.2011.06.002.

Dmitry Turaev. Polynomial approximations of symplectic dynamics and richness of chaos in non-hyperbolic area-preserving maps. *Nonlinearity*, 16(1):123–135, 2003. doi: 10.1088/0951-7715/16/1/308.

Mats Vermeeren, Alessandro Bravetti, and Marcello Seri. Contact variational integrators. *Journal of Physics A: Mathematical and Theoretical*, 52(44):445206, 2019. doi: 10.1088/1751-8121/ab4767.

Zhen Zhang, Yeonjong Shin, and George Em Karniadakis. GFINNs: GENERIC formalism informed neural networks for deterministic and stochastic dynamical systems. *Philosophical Transactions of the Royal Society A*, 380(2229):20210207, 2022. doi: 10.1098/rsta.2021.0207.

A Contact Geometry and Sign Conventions

Coordinates and Hamiltonian sign. We use $J^1(\mathbb{R}^n) \simeq \mathbb{R}^n_x \times \mathbb{R}_u \times \mathbb{R}^n_p$ with $\alpha = du - p \cdot dx$, taking p to be a row vector when derivatives of maps are written as matrices. The first jet of a smooth function v is $j^1v(x) = (x, v(x), \nabla v(x))$, and $(j^1v)^*\alpha = 0$. The Reeb vector field is $R_\alpha = \partial_u$. Our conventions define the contact Hamiltonian vector field X_H by

$$\alpha(X_H) = -H, \qquad \iota_{X_H} d\alpha = dH - (R_\alpha H)\alpha. \tag{10}$$

Since $d\alpha = \sum_i dx^i \wedge dp_i$, this gives precisely the coordinate equations in Equation (1). Cartan's formula then yields

$$\mathcal{L}_{X_H}\alpha = -(\partial_u H)\alpha, \qquad (\Phi^t_H)^*\alpha = e^{\lambda_\Phi(t,z)}\alpha, \qquad \lambda_\Phi(t,z) = -\int_0^t \partial_u H(\Phi^\tau_H(z))\, d\tau. \tag{11}$$

In particular, Φ^t_H is a contact map. For an autonomous H, the same equations give $\dot{H} = -H\,\partial_u H$. Hence the flow is strict when $H_u = 0$, while $H_u = \gamma$ gives the factor $e^{-\gamma t}$ used in the oscillator benchmark. These identities fix the signs of both the momentum friction term and the conformal factor (Geiges, 2008; de León & Lainz Valcázar, 2019).

The two flow classes. When $H = H(x, p)$, the (x, p) equations close and are Hamilton's equations for the symplectic form $\sum_i dx^i \wedge dp_i$. The remaining equation, $\dot{u} = p \cdot \partial_p H - H$, transports the action along that symplectic trajectory. Thus, in this case, the flow is a strict lift of its (x, p) map. If instead

$$H(x, u, p) = f(x, u) + g(x, u) \cdot p, \tag{12}$$

then $\dot{x} = g(x, u)$ and $\dot{u} = -f(x, u)$ close on the (x, u) base, and the full contact flow is the first prolongation of that base flow (Olver, 1995; Kevrekidis, 2026). These two cases are, respectively, the strict and prolonged flow classes used by the local splitting theorem underlying Theorem 4.2.

Prolongation and its domain. For a smooth base diffeomorphism $B(x, u) = (X(x, u), U(x, u))$, let $A = X_x + X_u p$. On any domain where A is invertible, define

$$\mathrm{pr}(B)(x, u, p) = (X, U, P), \qquad P = (U_x + U_u p)A^{-1}. \tag{13}$$

Here the matrix representation of X_x is $n \times n$, X_u is $n \times 1$, and U_x and p are $1 \times n$. The defining identity $PA = U_x + U_u p$ gives

$$\mathrm{pr}(B)^*\alpha = dU - P\,dX = c_B(du - p \cdot dx), \qquad c_B = U_u - PX_u = \frac{\det DB}{\det A}. \tag{14}$$

Thus c_B is nonzero when B is a diffeomorphism and A is nonsingular. On the domains considered here, we take $c_B > 0$. The full map can remain smooth even at points where its (x, u) projection folds, but this particular prolongation chart requires A to stay invertible. The results here use compact subsets of $J^1(\mathbb{R}^n)$ on which the relevant factors are defined. They do not require a global prolongation chart or address topology outside this jet space.

B Proof of Strict LegendreNet Universality

Here we prove Theorem 4.1. The proof requires several lemmas. Write $\theta = p \cdot dx$ on $T^*\mathbb{R}^n \simeq \mathbb{R}^{2n}$, so that $\alpha = du - \theta$. The first step identifies the action update of a strict map from its symplectic projection.

Proposition B.1 (Strict lifts)**.** *Every smooth symplectomorphism $\phi = (X, P)$ of $\mathbb{R}^{2n}$ has a strict contact lift*

$$\mathcal{L}(\phi)(x, u, p) = \big(X(x, p), u + A_\phi(x, p), P(x, p)\big), \qquad dA_\phi = \phi^*\theta - \theta. \tag{15}$$

The lift is unique up to a constant vertical translation $(x, u, p) \mapsto (x, u + c, p)$. Conversely, every smooth strict contactomorphism of $J^1(\mathbb{R}^n)$ has this form for a symplectomorphism ϕ.

Proof. Because $d\theta$ is the standard symplectic form and ϕ is symplectic, the fact that d commutes with pullbacks shows that the one-form $\phi^*\theta - \theta$ is closed. Because $\mathbb{R}^{2n}$ is simply connected, it

admits a global primitive A_ϕ, which is unique up to the addition of a global constant. Substitution into Equation (15) gives $\mathcal{L}(\phi)^*\alpha = d(u + A_\phi) - \phi^*\theta = \alpha$.

For the converse, let $\Phi = (X, U, P)$ be strict. A strict map preserves both α and $d\alpha$, so it preserves their Reeb field: $\Phi_*\partial_u = \partial_u$. Hence X and P do not depend on u, and $U = u + A(x, p)$ for some A. The pullback equation becomes $dA = \phi^*\theta - \theta$ for $\phi = (X, P)$. Taking exterior derivatives shows that ϕ is symplectic, with A_ϕ again determined up to a global constant. This proves the vertical-translation claim. □

Given symplectomorphisms ϕ_1 and ϕ_2 and choices of strict lifts $\mathcal{L}(\phi_1)$, $\mathcal{L}(\phi_2)$, and $\mathcal{L}(\phi_2 \circ \phi_1)$, the primitives are related by

$$A_{\phi_2\circ\phi_1} = A_{\phi_1} + A_{\phi_2} \circ \phi_1 + c \tag{16}$$

for some constant c. Thus strict lifting commutes with composition up to a vertical translation. For the Hénon-like map $h_{f,\eta}(x, p) = (p + \eta, -x + \nabla f(p))$, direct differentiation gives

$$h_{f,\eta}^*\theta - \theta = d\big(-\langle x, p\rangle + f(p)\big). \tag{17}$$

Its strict lift is therefore exactly the atom $L_{f,\eta,\zeta}$ in Equation (2), with ζ choosing the vertical constant.

The approximation argument also needs the following continuity result for strict lifts. In the statement, $\pi_{x,p} : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n}$ denotes projection onto the (x, p) coordinates.

Lemma B.2 (Continuity of strict lifts)**.** *Let $r \geq 1$ be an integer, let $K \subset J^1(\mathbb{R}^n)$ be compact, and choose a compact convex neighborhood $Q \subset \mathbb{R}^{2n}$ of $\pi_{x,p}(K)$. If $\{\phi_j\}$ is a sequence of symplectomorphisms such that $\phi_j \to \phi$ in $C^r(Q)$, then their strict lifts, normalized by $A_{\phi_j}(z_*) = A_\phi(z_*)$ at a common point $z_* \in Q$, satisfy $\mathcal{L}(\phi_j) \to \mathcal{L}(\phi)$ in $C^r(K)$. In particular, $C^1(Q)$ convergence of the symplectic maps also suffices for $C^0(K)$ convergence of their normalized lifts.*

Proof. Set $\eta_j = \phi_j^*\theta - \theta$ and $\eta = \phi^*\theta - \theta$. Each coefficient of these one-forms depends on the map and its first derivative, so $\|\eta_j - \eta\|_{C^{r-1}(Q)} \to 0$. Choose $z_* \in Q$ and normalize the primitives by $A_{\phi_j}(z_*) = A_\phi(z_*)$. The segment from z_* to any $z \in Q$ lies in Q. With $v = z - z_*$,

$$A_{\phi_j}(z) - A_\phi(z) = \int_0^1 \big[\iota_v(\eta_j - \eta)\big](z_* + tv)\, dt. \tag{18}$$

This formula bounds the $C^0(Q)$ norm of the primitive difference by $\operatorname{diam}(Q)\|\eta_j - \eta\|_{C^0(Q)}$. Moreover, $d(A_{\phi_j} - A_\phi) = \eta_j - \eta$, so every derivative of the primitive difference of order 1 through r is a derivative of a coefficient of $\eta_j - \eta$ of order at most $r - 1$. Thus the normalized primitives converge in $C^r(Q)$. The (x, p) components already converge, so the lifts converge in $C^r(K)$. The $C^0(K)$ conclusion under $C^1(Q)$ approximation follows from the case $r = 1$.

The estimate is locally linear in the approximation error. Indeed, writing $\phi_j = (X_j, P_j)$ and $\phi = (X, P)$ gives $\eta_j - \eta = (P_j - P) \cdot dX_j + P \cdot d(X_j - X)$. The product estimate in $C^{r-1}(Q)$ and the primitive bounds above imply

$$\|\mathcal{L}(\phi_j) - \mathcal{L}(\phi)\|_{C^r(K)} \leq C\|\phi_j - \phi\|_{C^r(Q)}, \qquad r \geq 1,$$

for the normalized lifts, where $C \geq 1$ depends on r, Q, and a uniform bound for the $C^r(Q)$ norms of the symplectic maps. Thus C can be fixed on a bounded $C^r(Q)$ neighborhood of the target. □

Proof of Theorem 4.1. By Proposition B.1, we may write the target as a strict lift of a symplectomorphism ϕ, with a chosen vertical constant. Let Q be a compact convex neighborhood of $\pi_{x,p}(K)$. Apply Turaev's Hénon-map approximation theorem (Turaev, 2003) at order $s = \max\{r, 1\}$ to obtain a finite composition

$$\widetilde{\phi} = h_{f_m,\eta_m} \circ \cdots \circ h_{f_1,\eta_1}, \qquad \|\widetilde{\phi} - \phi\|_{C^s(Q)} < \delta, \tag{19}$$

where the f_j are polynomials and δ can be arbitrarily small. By the estimate in Lemma B.2 at order s, choose $0 < \delta \leq \min\{1, \varepsilon/C\}$, with C fixed on the unit $C^s(Q)$ neighborhood of ϕ. Then a normalized strict lift of $\widetilde{\phi}$ is closer than ε to the target in $C^r(K)$, also when $r = 0$ since $s = 1$ in that case. Equations (16) and (17) identify that lift with a composition of Legendre atoms, up to one vertical translation. Assign that translation to one atom's ζ parameter. This gives the required LEGENDRENET. □

Remark B.3. The polynomial potentials f_j already suffice for this existence result. If the chosen trainable potential family instead approximates the finitely many f_j in C^{r+1} on their compact intermediate momentum domains, its atoms converge in C^r. Continuity of a finite composition then gives the same $C^r(K)$ conclusion for that family.

C Proof of Modular ContactNet Approximation

This appendix proves Theorem 4.2 and Corollary 4.3 using the local universal splitting theorem of Kevrekidis (2026).

Splitting theorem. For our purposes, the main result of Kevrekidis (2026) is the following. For a compact initial set K, a smooth contact Hamiltonian flow Φ_H^h defined over the time of interest, any finite r, and any $\delta > 0$, there exists a finite composition

$$S = \Theta_m \circ \cdots \circ \Theta_1, \qquad \|S - \Phi_H^h\|_{C^r(K)} < \delta, \tag{20}$$

whose factors Θ_i, for $i = 1, \ldots, m$, are exact strict flows or first prolongations of base flows. The construction assumes that these subflows and their finite compositions remain defined on compact neighborhoods of the relevant trajectories. The cited work proves this splitting result. Here we prove the theorems by showing that trainable approximations of its factors inherit the approximation claim.

Lemma C.1 (Stability of finite composition)**.** *For $m \geq 1$, let $F_m \circ \cdots \circ F_1$ be a composition of C^r maps defined along a compact chain $K_0 = K$ and $K_j = F_j(K_{j-1})$. Choose compact neighborhoods Q_j with $K_j \subset \operatorname{int} Q_j$ on which the next factor is defined. For every $\varepsilon > 0$, sufficiently small $C^r(Q_{j-1})$ errors in all factors produce a composition defined on K and within ε of the original composition in $C^r(K)$.*

Proof. We proceed by induction on m. For one factor the claim is immediate. Assume it holds for the first $m-1$ factors. A sufficiently small C^0 error keeps their perturbed image of K inside Q_{m-1}, so composition with the perturbed final factor is defined. Its error is the sum of the final-factor perturbation evaluated on that image and the change in F_m caused by perturbing its input. Uniform continuity controls the latter for $r = 0$. For higher derivatives, the chain rule expresses each derivative through order r as a finite sum of products of factor derivatives through order r. Uniform continuity of these derivatives on Q_{m-1} and the induction hypothesis make every term small. □

Lemma C.2 (One-derivative continuity of prolongation)**.** *Let $B = (X, U)$ be a C^{r+1} base diffeomorphism and let $Q \subset J^1(\mathbb{R}^n)$ be compact. Suppose $\sigma = X_x + X_u p$ is invertible on Q. If $\{B_k\}$ is a sequence of base diffeomorphisms such that $B_k \to B$ in C^{r+1} on a compact neighborhood of $\pi_{x,u}(Q)$, then $\operatorname{pr}(B_k)$ is defined on Q for all large k and $\operatorname{pr}(B_k) \to \operatorname{pr}(B)$ in $C^r(Q)$. Moreover, if the conformal factor of $\operatorname{pr}(B)$ is positive on Q, the factors of $\operatorname{pr}(B_k)$ are positive there for all large k.*

Proof. Write $\nu = U_x + U_u p$ and $P = \nu\sigma^{-1}$ as in Equation (13). Compactness gives a uniform bound on $\|\sigma^{-1}\|_{C^0(Q)}$. By assumption, the matrices $\sigma_k = (X_k)_x + (X_k)_u p$ converge uniformly to σ, so a Neumann-series argument keeps σ_k invertible with uniformly bounded inverses for large k. Matrix inversion is smooth on this bounded set of invertible matrices. Since the first derivatives of B_k converge in C^r, the product $\nu_k \sigma_k^{-1}$ converges in $C^r(Q)$, as do the base components. The conformal factor is the continuous expression $U_u - PX_u$ from Equation (14). A positive factor has a positive minimum on Q, which persists under close approximation. □

The derivative requirement cannot be replaced by value-level approximation, as demonstrated by the following example. In one dimension, consider $B_k(x, u) = (x + \sin(kx)/(2k), u)$. Each B_k is a base diffeomorphism because its spatial derivative is at least $1/2$, and $B_k \to \mathrm{id}$ uniformly. Yet its prolonged momentum is $P = p/(1 + \frac{1}{2}\cos(kx))$, which does not converge uniformly to p on any domain containing a nonzero momentum and a nontrivial spatial interval.

Proof of Theorem 4.2. Fix r and $\varepsilon > 0$. By the local universal splitting theorem of Kevrekidis (2026), there exists a composition S as in Equation (20) with error below $\varepsilon/2$ in $C^r(K)$. The compact-domain assumption gives a chain of compact neighborhoods for its factors. By Lemma C.1, there is a

factor tolerance $\eta > 0$ such that replacing every factor Θ_j by a C^r approximation within η keeps the full composition defined and changes it by less than $\varepsilon/2$ on K.

For strict and prolonged factors, we use the density hypotheses on $\mathcal{A}$ and $\mathcal{B}$, respectively. For a strict factor, cut off its u-independent Hamiltonian outside the compact trajectory tube to obtain a global strict flow that agrees with the original factor on the chosen compact neighborhood. Apply the density hypothesis on $\mathcal{A}$ there. For a prolonged factor, write $\Theta_j = \mathrm{pr}(b_j)$ with b_j a base-flow map. Its prolongation denominator is uniformly nonsingular on the chosen compact input neighborhood. Approximate b_j there in C^{r+1} by a member of $\mathcal{B}$. By Lemma C.2, its prolongation is within η of Θ_j in C^r and remains nonsingular. If consecutive factors have the same type, insert an exact identity factor of the other type before approximation. The density hypotheses approximate these identities as closely as needed, giving an alternating finite composition. Lemma C.1 and the triangle inequality now give an error below ε on K. □

Lemma C.3 (Neural-flow base approximators)**.** *Let $k \geq 1$. Exact flow maps of smooth time-dependent neural vector fields on $\mathbb{R}^{n+1}$ with tanh activation form a k-approximator for smooth base-flow maps, on compact sets whose trajectories exist throughout the prescribed time interval.*

Proof. Choose a compact tube with positive clearance around the target trajectories and extend the target vector field smoothly outside it. Tanh networks approximate this field, componentwise, in C^k on the compact time–space tube by Hornik (1991, Theorem 3). Gronwall's inequality first gives uniform convergence of trajectories; the clearance keeps the approximating trajectories in the tube. Differentiating the flow equation with respect to initial conditions gives the variational equations. At order j, the highest flow derivative enters linearly with coefficient $D_y v$, and the remaining terms use derivatives of v through order j and flow derivatives of lower order. Induction for $j = 1, \ldots, k$, uniform continuity of the top derivatives, and Gronwall give C^k convergence of the resulting map. Each exact neural flow is a diffeomorphism, with inverse obtained by reversing its time-dependent evolution. The statement concerns exact flows; a numerical solver introduces a separate discretization error. □

Augmented coupling alternative. Lyu et al. (2022, Theorems 3.2–3.6) approximate smooth single-coordinate diffeomorphisms in C^k by zero-padding, an augmented coupling network, and projection. A smooth base flow may be cut off outside its compact trajectory tube without changing its map there; decomposing the resulting compactly supported map and composing the projected approximants gives C^k approximation on enlarged compact neighborhoods. For sufficiently close C^1 approximants the Jacobians remain nonsingular and injectivity holds after shrinking to the required compact neighborhoods: nearby pairs are separated by local inverse estimates, and pairs a fixed distance apart by compactness and injectivity of the target. Thus these restrictions are locally invertible base maps on the neighborhoods required by the modular theorem. Global invertibility of an arbitrary projected network lies outside this construction, and its inverse is not the algebraic augmented inverse because the discarded coordinates are unknown.

Proof of Corollary 4.3. For (a), Theorem 4.1 supplies the strict family and Theorem 4.2 transfers C^{r+1} base density to C^r contact density. Lemma C.3 verifies the base hypothesis for the stated neural-flow instantiation. For (b), apply the same theorem under the explicitly stated triangular-family hypothesis; the algebraic inverse formulas in Appendix D hold independently of that hypothesis. Set optional vertical blocks to the identity. □

D Architecture Formulas and Inverses

This section gives the operations used to evaluate and invert the blocks in Section 3. All inverse formulas are evaluated at an output jet (X, U, P). Their stated domains matter when a prolonged map approaches a singular projection.

D.1 Strict Legendre atoms

For the atom $L_{f,\eta,\zeta}$ in Equation (2), the output position determines the old momentum. The complete algebraic inverse is

$$p = X - \eta, \qquad x = \nabla f(p) - P, \qquad u = U + \langle x, p\rangle - f(p) - \zeta. \tag{21}$$

No inverse of f or nonlinear solve is required. Direct differentiation of $U = u - \langle x, p\rangle + f(p) + \zeta$ gives $dU - P\cdot dX = du - p\cdot dx$, verifying strict contactness. When $f = 0$ and $\eta = \zeta = 0$, one atom is the Hénon swap $(x, u, p) \mapsto (p, u - \langle x, p\rangle, -x)$. Four such swaps compose to the identity. Zeroing the potential readout and both shifts therefore provides an identity initialization for a four-atom strict block.

D.2 One-dimensional triangular prolongations

Consider the representative base map used in the 1D architecture,

$$U = a(x)u + b(x), \qquad X = c(U)x + d(U), \qquad a(x) > 0, \quad c(U) > 0. \tag{22}$$

The positive scales make both triangular steps invertible. Write $q = c'(U)x + d'(U)$. The derivatives needed for its first prolongation are

$$U_x = a'(x)u + b'(x), \quad U_u = a(x), \quad X_x = c(U) + qU_x, \quad X_u = qU_u. \tag{23}$$

Consequently, with $\sigma = X_x + pX_u$, the forward momentum and contact factor are

$$P = \frac{U_x + pU_u}{\sigma}, \qquad c_B = U_u - PX_u = \frac{a(x)c(U)}{\sigma}. \tag{24}$$

The last equality follows from $\det DB = a(x)c(U)$. The base variables and old momentum are recovered in this order:

$$x = \frac{X - d(U)}{c(U)}, \qquad u = \frac{U - b(x)}{a(x)}, \qquad p = \frac{PX_x - U_x}{U_u - PX_u}, \tag{25}$$

where the derivatives in the final expression are evaluated at the recovered (x, u). The forward formula requires $\sigma \neq 0$, and the inverse formula requires $U_u - PX_u \neq 0$. On the image of a regular forward map, these conditions agree through $(U_u - PX_u)\sigma = a(x)c(U) > 0$. The triangular implementation leaves the forward denominator unclamped, so states near $\sigma = 0$ can expose an unstable projection. A sampled denominator penalty discourages such states but does not make the prolongation global.

The same one-dimensional prolongation formula applies if the order of the two triangular base steps is reversed. In that case the base inverse unwinds the action step before the spatial step. Neither order requires a separately learned inverse.

D.3 Higher-dimensional prolonged blocks

For $x, p \in \mathbb{R}^n$, a u-conditioned spatial block has

$$X = a(u)x + b(u), \qquad U = u, \qquad a(u) > 0, \quad b(u) \in \mathbb{R}^n. \tag{26}$$

Set $w = a'(u)x + b'(u) \in \mathbb{R}^n$. Then $X_x + X_u p = a(u)I + wp$ is a rank-one update of a scalar matrix. The Sherman–Morrison formula reduces the prolonged momentum to

$$P = \frac{p}{a(u) + p\cdot w}, \qquad x = \frac{X - b(U)}{a(U)}, \qquad p = \frac{a(U)P}{1 - P\cdot w}. \tag{27}$$

Here w in the inverse is evaluated at the recovered (x, U). The forward prolongation is defined when $a(u) + p\cdot w \neq 0$. On its image, $1 - P\cdot w = a(u)/(a(u) + p\cdot w) \neq 0$. Its contact factor is the same ratio $1 - P\cdot w$, which is required to be positive on the domains considered here.

A complementary x-conditioned action shear leaves the position fixed:

$$X = x, \qquad U = a(x)u + b(x), \qquad P = a(x)p + u\nabla a(x) + \nabla b(x), \qquad a(x) > 0. \tag{28}$$

Its inverse first sets $x = X$ and $u = (U - b(X))/a(X)$, then uses

$$p = \frac{P - u\nabla a(X) - \nabla b(X)}{a(X)}. \tag{29}$$

The factor of this block is $a(x) > 0$, and its prolongation has no momentum-dependent denominator. Compositions of these two base-map types admit algebraic inverses by reversing their order.

D.4 Residual action blocks

The optional vertical block in the 1D benchmark uses a smooth scalar field $F(x, u)$ and a scale ρ:

$$X = x, \qquad U = u + \rho F(x, u), \qquad P = \rho F_x(x, u) + (1 + \rho F_u(x, u))p. \tag{30}$$

Its contact factor is $1 + \rho F_u$. For a given (X, U, P), the implemented inverse starts at $u^{(0)} = U$ and iterates

$$u^{(k+1)} = U - \rho F(X, u^{(k)}), \qquad p = \frac{P - \rho F_x(X, u)}{1 + \rho F_u(X, u)} \tag{31}$$

after the scalar iteration. The reported 1D inverse uses 30 iterations. If $\sup |\rho F_u| < 1$ on an invariant interval containing these iterates, the scalar map is a contraction and the iteration converges to the unique inverse there. The weaker condition $1 + \rho F_u \neq 0$ ensures only local invertibility. The action-Jacobian penalty promotes a margin on sampled states but does not imply the contraction bound.

D.5 Composition and derivative evaluation

For one stage $V_j \circ C_j \circ L_j$, the forward pass evaluates the strict block, prolonged block, and optional vertical block in that order. The inverse uses $L_j^{-1} \circ C_j^{-1} \circ V_j^{-1}$, with the inverse stages then evaluated from last to first. The same fitted parameters appear in both directions. Strict atoms and triangular prolongations require values and first derivatives of their scalar fields. For these blocks in JAX, shallow tanh fields return these derivatives from explicit formulas, while deeper fields use automatic differentiation. The contact identities above hold for either choice of smooth field. Numerical diagnostics remain subject to floating-point error. Their precision depends on the experiment.

E Experimental Setup and Reproducibility

Table 1: Benchmark suite. The protocols differ and are not pooled into a single ranking.

Study	Evaluation role	Detailed results
Strict lifting	Separate phase accuracy from the compatible action update; measure rollout, inverse, and strict-contact defect	Appendix F.1
Oscillators	Compare triangular and neural-flow bases, one-step and long-rollout error, inversion and contact defect	Appendix F
Action dependence	Vary action power, depth, and capacity; compare full-state rollout with the response to changing initial action	Appendices F.3 and F.3.1
Fronts and caustics	Measure state and Legendrian error, data efficiency, and inverse recovery in one and two dimensions	Section 7 and its appendices

E.1 Benchmark data and splits

Table 2 summarizes the primary long-rollout datasets used in the main text. Controlled lifting, factorization, and ablation studies state their protocol changes alongside their results. All fitted models learn a one-step map from complete jets and are evaluated by repeated application of that map. The 1D and 2D front rollouts stay within the time interval represented in training. Only the oscillator rollout from π to 2π tests a later time interval.

Table 2: Primary reference-data protocols. Points count spatial or ray labels per initial condition. Adjacent states at saved training times form the one-step training pairs.

Dataset	Dimension	Points	Times	Δt	Train / val. / test	Reference solver
Oscillator	1	1	81	$\pi/40$	64 / 16 / 32	DOP853
1D fronts	1	64	13	0.1	40 / 0 / 15	RK4, 16 substeps
2D fronts	2	20^2	13	0.1	30 / 0 / 11	RK4, 16 substeps

E.1.1 DYNAMICAL-SYSTEM DATA

For the oscillator, the data generator samples x_0, p_0 uniformly from $[-1, 1]$ and u_0 uniformly from $[-0.5, 0.5]$ using data seed 20260917. The 64 training trajectories supply their first 40 transitions, covering $0 \le t \le \pi$. Validation uses 16 disjoint trajectories over the same interval. Standard testing uses another 32 trajectories and runs for 80 steps to 2π. The DOP853 reference solve uses relative tolerance 10^{-12} and absolute tolerance 10^{-13}. Independent checks compare (x, p) with its matrix-exponential solution and the effect of an action shift with $e^{-\gamma t}$. All oscillator arrays and model calculations use double precision.

E.1.2 HAMILTON–JACOBI AND LEGENDRIAN-FRONT DATA

The 1D datasets use the same 55 sampled initial profiles across the three Hamiltonians in Equation (33). A profile is drawn with equal probability from a Gaussian or cosine family, with wavefront seed zero. Gaussian amplitude lies in $[-1.5, 1.5]$, width in $[0.4, 0.9]$, and center in $[-0.5, 0.5]$. Cosine amplitude lies in $[0.5, 1.5]$, wave number in $[0.8, 2.2]$, and phase in $[0, 2\pi]$. The labels are 64 equally spaced points on $[-1.5, 1.5]$. Forty profiles form training pairs and 15 are held out. Each reference trajectory is saved from $t = 0$ to 1.2 at spacing 0.1, with 16 RK4 substeps between saved times. The exported training and trajectory arrays are float32, with no added observation noise. The stored intensity dataset does not include β; the generator setting is $\beta = 0.5$.

The 2D benchmark uses $H = \|p\|^2/2 + u^2/4$. Its 30 training profiles mix Gaussian bumps and astigmatic profiles with equal sampling probability. The ten sampled test profiles are disjoint from training, and an additional fixed Gaussian profile brings the test total to eleven. Every profile is evaluated on a 20×20 label grid over $[-1, 1]^2$ and integrated to $t = 1.2$ at spacing 0.1, again with 16 RK4 substeps per saved interval. The experiment uses wavefront seed zero and one model initialization seed, zero. Its inverse illustration uses one held-out terminal jet with supplied momentum.

E.2 MODEL CONFIGURATIONS AND OPTIMIZATION

E.2.1 DYNAMICAL-SYSTEM MODELS

The oscillator architecture sizes are shown in Table 3. A pair means one strict Legendre block followed by one prolonged base block when enabled. Atoms count Hénon-like maps within a strict block. The LEGENDRENET rows omit the prolonged block. GCF uses its own splitting blocks and random Fourier features, while RealNVP uses affine coupling layers on the full jet. Parameter counts include trainable scalars but exclude fixed random features.

Neural-flow base. CONTACTNET-ODE retains the strict blocks and replaces the triangular base and optional vertical block with a neural ODE on $y = (x, u)$. Its time-dependent field $f_\theta(\tau, y)$ is a tanh MLP with two width-20 hidden layers. Diffrax 0.7.2 (Kidger, 2021) integrates $\dot{y} = f_\theta(\tau, y)$ and $\dot{J} = D_y f_\theta(\tau, y) J$, $J(0) = I_2$, together over $[0, 1]$ using four fixed Tsit5 steps (Tsitouras, 2011). Autodiff evaluates $D_y f_\theta$; the propagated Jacobian supplies the derivatives in Equation (4). Training differentiates through the solve. The exact-flow family is covered by Lemma C.3; the solver introduces a separate discretization error. Both variants use the same data protocols and optimization schedule. Unqualified CONTACTNET denotes the triangular variant in the tables below.

Table 3: Oscillator models. Each row uses three initialization seeds, 80 Adam epochs, and 40 full-batch L-BFGS steps. All use learning rate 2×10^{-3}, batch size 256, and gradient clipping at 1.

Model	Structure	Width / features	Parameters	Extra setting
MLP	3 layers	48	5,043	residual
RealNVP	6 layers	12	2,460	depth 2
LEGENDRENET	4 blocks / 4 atoms	48	2,352	no base block
GCF	8 blocks	120 features	2,880	$\tau = 0.2$
CONTACTNET	3 pairs / 4 atoms	32	2,739	x, u coupling; vertical block
CONTACTNET-ODE	3 pairs / 4 atoms	32 / 20	2,814	neural ODE; no vertical block

Oscillator models start from small random weights rather than the optional identity initialization. Both CONTACTNET variants receive a denominator penalty with weight one and margin 0.2; the

triangular variant also receives a vertical-Jacobian penalty with the same settings. The optimizer checkpoints at initialization and every ten Adam epochs or L-BFGS steps, retaining the lowest validation one-step MSE. The structural diagnostic uses autodiff Jacobians on a separate fixed cloud of 32 states. Reference-terminal inversion uses 80 applications of the frozen inverse. A separate 16-state self-cycle probe tests inversion of the model's own output. The neural-ODE runs evaluate forward maps only; inverse errors and inverse costs are not measured.

E.2.2 HAMILTON–JACOBI AND LEGENDRIAN-FRONT MODELS

The one-dimensional front models use the same block terminology and are listed in Table 4.

Table 4: One-dimensional front models. All rows use 150 Adam epochs at learning rate 2×10^{-3}, batch size 1024, and gradient clipping at 1. L-BFGS is the number of reported steps.

Model	Structure	Width / features	Parameters	L-BFGS
MLP	3 layers	144	42,771	300
RealNVP	9 layers	25	13,518	500
LEGENDRENET	7 blocks / 4 atoms	160	13,524	300
GCF	12 blocks	375 features	13,500	300
CONTACTNET	5 pairs / 4 atoms	96	13,525	300

These 1D runs use initialization seeds $0, 1, 2$ and identity initialization where supported. The CONTACTNET configuration uses triangular x, u base coupling and the residual vertical block. Its denominator and action-Jacobian hinge penalties both have weight one and margin 0.2. The other rows receive no such penalties. The selected checkpoints minimize held-out rollout state error on the same 15 profiles used for reporting. Thus model selection and reporting use the same profiles. Keeping the last logged checkpoint instead leaves every ranking in Table 16 unchanged (Table 17). The MLP has about three times the parameter count of each invertible model, and RealNVP has more L-BFGS steps. The three structure-preserving rows share the nominal optimization schedule, though their training costs differ. The separate folded-front illustration uses a CONTACTNET with the same 5-pair, 4-atom, width-96 architecture, trained for 300 Adam epochs and 2,000 L-BFGS steps. The same held-out rule kept its checkpoint at L-BFGS step 1,700, with rollout state MSE 2.02×10^{-5} against 2.06×10^{-5} at step 2,000. Its outcomes are single-model results.

The 2D experiment uses a width-96 CONTACTNET with ten strict and prolonged stages, two Legendre layers and six atoms per Legendre layer, independent atom parameters, and u-conditioned spatial coupling. Its denominator hinge has weight one and margin 0.2. The MLP has three layers of width 96. The strict comparison uses ten symplectic layers of width 96 and reports only (x, p) error. All three models use one-step AdamW for 120 epochs at learning rate 2×10^{-3}, gradient clipping at 1, and a spatial-grid Jacobian loss with weight 3. The contact model also attempted 500 full-batch L-BFGS steps. Its selected checkpoint was the AdamW epoch-120 model because the later phase did not improve the held-out rollout score. The MLP and strict model had no L-BFGS phase. These 2D checkpoints were also chosen using the reported test profiles, and the results have no training-seed error bars.

E.3 BASELINE IMPLEMENTATION, METRICS, AND COMPUTING ENVIRONMENT

At the time of writing, the official project page for Testa et al. (2025)[1] described its code as forthcoming; no author-released implementation was available for these experiments. We therefore implemented the proposed GCF architecture from the published equations and reported settings to the best of our ability. Every GCF result in this paper concerns that independent implementation.

State MSE averages squared coordinate errors over the reported test states and rollout times. The Legendrian metric uses finite differences along sampled front labels and therefore has a nonzero discretization floor even for exact contact maps. The seeded and sparse studies both use physical label derivatives and mean squared residuals, not arclength or RMS. They use different profile generators despite the same Hamiltonian and 64-ray grid size; the seeded metric also includes $t = 0$, while the

[1] https://sites.google.com/view/geometric-contact-flows, accessed September 23, 2026.

sparse metric averages the 12 predicted times. Their floors are therefore dataset-dependent and not directly comparable. The reference residual is a discretization baseline, not a lower bound: a learned curve can have smaller finite-difference truncation error, explaining the two-front value slightly below the sparse reference. The oscillator contact defect instead uses a Jacobian pullback check on its independent state cloud. Inverse recovery starts from a reference terminal state or lifted reference front. For the 1D benchmark, the reported recovery MSE sums the three coordinate MSEs averaged over all thirteen states of the backward trajectory. It is not an error measured only at the recovered initial state. A self-cycle starts from a model state and applies that same model forward and backward. These quantities are reported separately.

The oscillator experiments used a CPU JAX backend on macOS and record Python and library versions, per-seed training seconds, selected checkpoints, and per-time errors. The 2D experiments used the CPU Torch backend. Wall time was recorded for the 1D front experiments, but the hardware specification was not retained. Wall times across these three experiments are therefore not directly comparable.

E.4 Computational cost and structural diagnostics

Table 5 gives the oscillator timings on the archived CPU JAX environment. Training seconds include the common 80-epoch Adam and 40-step L-BFGS schedule, but costs differ substantially with model evaluation and differentiation. Step timings are warmed, synchronous evaluations on a batch of 32 states. The GCF timing probe is incomplete across seeds because its wide rollouts failed, so we omit its per-step aggregate rather than reporting the one available seed as a three-seed comparison. The archives do not provide a common peak-memory measurement. The neural-ODE archive records Python 3.12.14, JAX 0.11.1, Equinox 0.13.8, Optax 0.2.8, and Diffrax 0.7.2, but no hardware identifier or per-step timing. Its training times (marked †) are reported for context rather than as a hardware-matched speed comparison.

Table 5: Oscillator computational cost and one-step contact diagnostics. Entries are three-seed mean ± population standard deviation, except contact defect, which is shown as a mean. Training is in seconds; forward and inverse costs are milliseconds per batch step. Dashes mark unavailable measurements or the MLP's absent inverse; † marks neural-ODE timings without recorded hardware.

Regime	Model	Train (s)	Forward (ms)	Inverse (ms)	Contact defect
Cons.	MLP	2.6 ± 0.1	0.087 ± 0.001	–	$6.82e{-3}$
Cons.	RealNVP	10.8 ± 0.1	0.182 ± 0.004	0.180 ± 0.003	$1.50e{-2}$
Cons.	LegendreNet	24.6 ± 0.2	0.212 ± 0.002	0.206 ± 0.012	$2.48e{-16}$
Cons.	GCF	145.0 ± 23.3	–	–	$1.84e{-16}$
Cons.	ContactNet	66.0 ± 12.1	0.311 ± 0.009	0.756 ± 0.095	$2.88e{-16}$
Cons.	ContactNet-ODE †	283.9 ± 3.5	–	–	$2.96e{-16}$
Damped	MLP	3.4 ± 0.3	0.093 ± 0.002	–	$7.61e{-3}$
Damped	RealNVP	12.3 ± 0.2	0.209 ± 0.024	0.205 ± 0.009	$1.90e{-2}$
Damped	LegendreNet	36.9 ± 5.7	0.434 ± 0.300	0.267 ± 0.046	$2.60e{-16}$
Damped	GCF	114.8 ± 17.2	–	–	$1.55e{-16}$
Damped	ContactNet	82.1 ± 22.7	0.516 ± 0.295	0.812 ± 0.154	$2.61e{-16}$
Damped	ContactNet-ODE †	287.2 ± 5.3	–	–	$3.14e{-16}$

F Controlled Dynamical-System Benchmarks

F.1 Strict lifting beyond Legendre blocks

The targets are strict free propagation and the unit harmonic oscillator, with analytic maps at step size $h = 0.15$ and initial jets uniform on $[-1, 1]^3$. For each seed, the 32 training points are a prefix of the 256-point set; independent validation and test sets each contain 256 points. All fits use 120 Adam epochs at learning rate 0.003, batches of at most 64, and minimum validation full-jet one-step MSE for selection. The 16-step rollout starts from the test jets. Seeds are 0, 1, and 2.

The Hénon backbone has two independent layers of four atoms, width 12. The G-SympNet has eight alternating gradient shears, width 12, following Jin et al. (2020). Both admit exact action cocycles.

For $X = x + V'(p)$, $P = p$, the action update is $U = u + pV'(p) - V(p)$; for $X = x$, $P = p + V'(x)$, it is $U = u + V(x)$. The learned-head variants instead use $U = u + f(x, u, p)$. Nominal parameter counts are 312/353 for exact/learned Hénon models and 304/345 for G-SympNet. Sixteen action-only leaves in the learned-head Hénon variant are unused, giving an effective count of 337. The learned action head is inverted with 12 Newton iterations; its inverse accuracy is measured rather than guaranteed globally.

The strict-contact defect is the RMS of the three coefficients of $dU - P\,dX - (du - p\,dx)$ on 64 independent probe points. Before fitting, the runner checks the component and composed maps, inverses, action shifts, and analytic references; oscillator action is additionally checked by independent quadrature. The recorded maximum exact-formula defect is 2.22×10^{-16} against a 10^{-9} failure threshold. All 48 fits passed the recorded finite-result check.

Table 6: Complete strict-lift comparison: three-seed means. Phase and action are one-step MSE; rollout averages full-jet MSE over 16 steps; inverse starts at a reference terminal jet. Defect is coefficient RMS.

Flow	N	Backbone/action	Phase	Action	Rollout	Defect	Inverse
Free	32	LEGENDRENET exact	$2.75e{-5}$	$3.63e{-5}$	$2.88e{-3}$	$8.87e{-17}$	$3.04e{-5}$
Free	32	LEGENDRENET learned	$7.88e{-6}$	$6.05e{-4}$	$1.91e{-2}$	$4.24e{-2}$	$2.11e{-4}$
Free	32	G-SympNet exact	$2.00e{-7}$	$2.22e{-7}$	$5.32e{-5}$	$4.57e{-17}$	$2.07e{-7}$
Free	32	G-SympNet learned	$8.66e{-6}$	$5.77e{-4}$	$1.86e{-2}$	$4.92e{-2}$	$2.00e{-4}$
Free	256	LEGENDRENET exact	$1.68e{-5}$	$2.16e{-5}$	$1.72e{-3}$	$8.55e{-17}$	$1.84e{-5}$
Free	256	LEGENDRENET learned	$7.65e{-6}$	$4.13e{-4}$	$1.34e{-2}$	$3.51e{-2}$	$1.42e{-4}$
Free	256	G-SympNet exact	$1.21e{-7}$	$1.39e{-7}$	$1.95e{-5}$	$4.83e{-17}$	$1.28e{-7}$
Free	256	G-SympNet learned	$1.37e{-7}$	$4.13e{-4}$	$1.29e{-2}$	$4.40e{-2}$	$1.37e{-4}$
Oscillator	32	LEGENDRENET exact	$6.62e{-4}$	$1.15e{-4}$	$5.70e{-2}$	$9.78e{-17}$	$4.78e{-4}$
Oscillator	32	LEGENDRENET learned	$2.24e{-5}$	$8.36e{-4}$	$1.96e{-2}$	$5.38e{-2}$	$2.99e{-4}$
Oscillator	32	G-SympNet exact	$1.42e{-6}$	$2.46e{-7}$	$5.08e{-5}$	$4.53e{-17}$	$1.02e{-6}$
Oscillator	32	G-SympNet learned	$4.36e{-6}$	$8.26e{-4}$	$1.82e{-2}$	$5.90e{-2}$	$2.84e{-4}$
Oscillator	256	LEGENDRENET exact	$5.23e{-5}$	$3.30e{-5}$	$6.91e{-3}$	$9.90e{-17}$	$4.48e{-5}$
Oscillator	256	LEGENDRENET learned	$1.55e{-5}$	$6.10e{-4}$	$1.48e{-2}$	$4.72e{-2}$	$2.12e{-4}$
Oscillator	256	G-SympNet exact	$1.66e{-7}$	$1.66e{-7}$	$1.56e{-5}$	$3.91e{-17}$	$1.62e{-7}$
Oscillator	256	G-SympNet learned	$1.90e{-7}$	$6.10e{-4}$	$1.41e{-2}$	$5.51e{-2}$	$2.03e{-4}$

These results isolate the exact action update within each phase architecture. Both backbones benefit in action accuracy and contact fidelity. The exact G-SympNet is particularly well aligned with the linear reference phase maps. The 32-point oscillator Hénon exception shows that lower action error need not imply lower full-jet rollout error. Seed variation here measures initialization variability for the specified data protocol, not uncertainty across physical systems.

F.2 OSCILLATOR ONE-STEP, ROLLOUT, AND INVERSE ACCURACY

Table 7 separates the one-step fit from errors at the end of the training interval and after 80 autoregressive steps. Entries are three-seed means and population standard deviations. Reference-state inversion starts at the exact terminal state and applies the frozen inverse for 80 steps; it is distinct from a model self-cycle. The contact defect is the mean norm of the Jacobian pullback residual on a fixed independent cloud.

Table 7: Oscillator standard-test errors: three-seed mean $\pm$ population standard deviation. Contact defect is shown as a three-seed mean. The MLP has no implemented inverse; neural-ODE inversion was not evaluated.

Regime	Model	One-step MSE	$t = \pi$ MSE	$t = 2\pi$ MSE	Inverse MSE	Contact defect
Cons.	MLP	$5.01e{-6} \pm 7.00e{-8}$	$2.19e{-3} \pm 5.13e{-4}$	$7.05e{-3} \pm 9.77e{-4}$	–	$6.82e{-3}$
Cons.	RealNVP	$1.63e{-5} \pm 4.72e{-6}$	$7.11e{-3} \pm 2.13e{-3}$	$2.10e{-2} \pm 2.07e{-2}$	$2.51e{-2} \pm 2.40e{-2}$	$1.50e{-2}$
Cons.	LEGENDRENET	$6.38e{-6} \pm 6.15e{-7}$	$5.88e{-3} \pm 4.54e{-4}$	$2.17e{-2} \pm 1.79e{-3}$	$2.81e{-2} \pm 3.23e{-3}$	$2.48e{-16}$
Cons.	GCF	$3.38e{-6} \pm 9.18e{-7}$	$1.51e{-3} \pm 7.06e{-4}$	$3.62e{-3} \pm 1.85e{-3}$	$8.90e{-3} \pm 4.94e{-3}$	$1.84e{-16}$
Cons.	CONTACTNET	$1.35e{-5} \pm 3.57e{-6}$	$1.37e{-2} \pm 5.29e{-3}$	$4.75e{-2} \pm 1.10e{-2}$	$2.31e{-2} \pm 5.01e{-3}$	$2.88e{-16}$
Cons.	CONTACTNET-ODE	$1.91e{-5} \pm 4.15e{-6}$	$2.73e{-2} \pm 4.76e{-3}$	$9.08e{-2} \pm 1.30e{-2}$	–	$2.96e{-16}$
Damped	MLP	$3.17e{-6} \pm 3.45e{-7}$	$1.44e{-3} \pm 1.90e{-4}$	$2.14e{-3} \pm 3.12e{-4}$	–	$7.61e{-3}$
Damped	RealNVP	$1.31e{-5} \pm 3.25e{-6}$	$5.36e{-3} \pm 1.98e{-3}$	$3.96e{-3} \pm 2.15e{-3}$	$1.98e{-2} \pm 1.84e{-2}$	$1.90e{-2}$
Damped	LEGENDRENET	$1.97e{-5} \pm 2.97e{-7}$	$2.28e{-2} \pm 3.85e{-3}$	$7.12e{-2} \pm 1.04e{-2}$	$5.39e{-2} \pm 1.61e{-3}$	$2.60e{-16}$
Damped	GCF	$1.89e{-6} \pm 7.46e{-7}$	$2.87e{-4} \pm 1.50e{-4}$	$2.27e{-4} \pm 9.04e{-5}$	$9.43e{-4} \pm 4.85e{-4}$	$1.55e{-16}$
Damped	CONTACTNET	$1.24e{-5} \pm 6.47e{-6}$	$1.13e{-2} \pm 7.25e{-3}$	$1.09e{-2} \pm 4.44e{-3}$	$1.56e{-2} \pm 6.25e{-3}$	$2.61e{-16}$
Damped	CONTACTNET-ODE	$1.40e{-5} \pm 5.68e{-6}$	$1.35e{-2} \pm 1.89e{-3}$	$1.84e{-2} \pm 3.46e{-3}$	–	$3.14e{-16}$

The one-step ordering is not a reliable guide to the 80-step ordering. For example, the damped LEGENDRENET one-step error is of order 10^{-5}, while its mean final-time error is 7.12×10^{-2}. GCF has the lowest standard-test mean final-time error in both regimes. The contact architectures have pullback defects at numerical precision, while their trajectory errors remain target and architecture dependent. CONTACTNET-ODE has higher final-time error than the triangular variant in both regimes and all three seeds.

Table 8 reports each wider-range rollout separately. The wide initial intervals are twice those used for training, while the reference solver and terminal time are unchanged. The marker F_k means that nonfinite values first appeared at step k; the corresponding final-time MSE is unavailable. GCF has only one complete wide rollout per regime. No three-seed wide-range GCF mean is reported. CONTACTNET completes all six wide rollouts, but its conservative seed 2 reaches 8.93×10^{33} MSE. Finite output alone is therefore not evidence of accurate extrapolation.

Table 8: Wider-range oscillator final-time MSE by training seed. F_k marks the first nonfinite step of an incomplete rollout.

Regime	Model	Seed 0	Seed 1	Seed 2
Cons.	MLP	$5.25e{-1}$	$6.80e{-1}$	$4.45e{-1}$
Cons.	RealNVP	$7.25e{+2}$	$4.98e{+1}$	$4.36e{+0}$
Cons.	LEGENDRENET	$2.27e{+0}$	$2.27e{+0}$	$2.85e{+0}$
Cons.	GCF	F_{67}	$1.61e{+3}$	F_{58}
Cons.	CONTACTNET	$1.27e{+0}$	$1.83e{+0}$	$8.93e{+33}$
Cons.	CONTACTNET-ODE	$5.04e0$	$5.62e0$	$6.10e1$
Damped	MLP	$1.52e{-1}$	$2.07e{-1}$	$2.50e{-1}$
Damped	RealNVP	$1.93e{+13}$	$1.62e{+0}$	$3.24e{+0}$
Damped	LEGENDRENET	$2.09e{+0}$	$2.14e{+0}$	$3.16e{+0}$
Damped	GCF	F_{61}	$1.49e{+0}$	F_{79}
Damped	CONTACTNET	$3.68e{-1}$	$3.08e{-1}$	$3.48e{-1}$
Damped	CONTACTNET-ODE	$1.86e0$	$2.23e0$	$7.37e1$

F.3 CONTROLLED FACTORIZATION AND ACTION TRANSPORT

The depth study fits $H = p^2/2 + x^2/2 + \gamma_k u^k/k$ for $k = 1, 2, 3, 4$. The same sampled profile families, grids, and independent train/validation/test indices are used across powers. The coefficient γ_k calibrates the RMS initial dissipative momentum force to 0.12. This controls a scale without equating the difficulty of the targets. All fits use 400 Adam updates and validation-only checkpoint selection, with three seeds and depths 1, 2, 4, and 8. Depth counts GCF blocks or CONTACTNET strict/prolonged pairs and is architecture dependent. The latter uses four shared Legendre atoms and one triangular coupling per pair, without the optional vertical action block.

The fixed-per-block view uses width 16 for CONTACTNET and 80 Fourier features for GCF. The approximately fixed-total view uses width $\max(4, \mathrm{round}(64/d))$ and five times as many GCF features at depth d. All 192 attempts completed without failure. The maximum recorded differential contact defect was 9.30×10^{-16}. Table 9 gives a representative matched-capacity slice, while Figure 5 summarizes all powers and depths in both capacity views. Positive heatmap entries favor CONTACTNET and negative entries favor GCF; zero denotes equal median one-step error.

Table 9: Representative depth-four results at approximately fixed total capacity. Entries are median one-step MSE over three seeds; the lower value in each row is bold.

Action power	GCF	CONTACTNET	Lower median
1	$\mathbf{2.47e{-7}}$	$7.50e{-6}$	GCF
2	$6.75e{-6}$	$\mathbf{2.78e{-6}}$	CONTACTNET
3	$\mathbf{3.30e{-6}}$	$4.68e{-6}$	GCF
4	$4.72e{-6}$	$\mathbf{3.23e{-6}}$	CONTACTNET

At depth four with approximately fixed total capacity, GCF has lower median one-step error for powers one and three, and CONTACTNET for powers two and four. The ranking varies at other

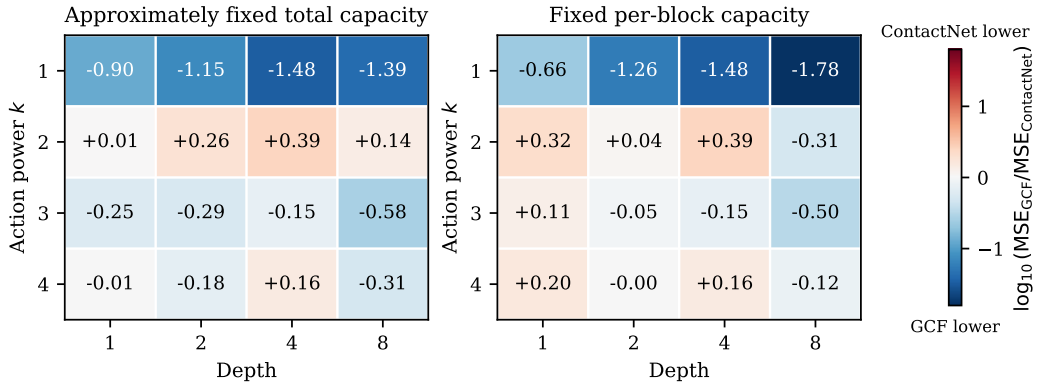

Figure 5: Finite-depth factorization comparison over all 32 power/depth/capacity combinations. Cells show $\log_{10}(\mathrm{MSE}_{\mathrm{GCF}}/\mathrm{MSE}_{\mathrm{ContactNet}})$ for the median one-step error over three seeds. Red favors ContactNet, blue favors GCF.

depths. The accompanying symbolic calculation shows that sums and first contact brackets of the implemented GCF generator class remain at most linear in u. This characterizes the generator algebra through first brackets; trained-error bounds and approximation rates for finite compositions are separate questions. At the level of generator classes, $T(p)+V(x)$ is strict and $F(x)u$ is prolonged, so the modular theorem encompasses those generator types. Containment and performance comparisons between fixed trained architectures remain separate finite-capacity questions.

F.3.1 Action-dependent position transport

The companion targets are $H = \epsilon up$ and $H = \epsilon \tanh(u)p$. They use the oscillator protocol: 64 training, 16 validation, and 32 test initial states, three seeds, 80 Adam epochs followed by 40 L-BFGS steps, and 80 rollout steps to 2π. A separate held-out probe fixes initial (x, p) and compares final position for $u = -0.4$ and $u = 0.4$. The exact final contrasts are 0.17593 and 0.16711, respectively. The table distinguishes error in this contrast from full-state trajectory error.

Table 10: Action transport at $t = 2\pi$: three-seed means of standard-test and wider-range full-state MSE, and held-out paired-action position-contrast MSE. GCF has only one complete wide rollout per target, so its wide mean is omitted.

Target	Model	Test state MSE	Contrast MSE	Wide state MSE
Linear	MLP	$1.94e{-3}$	$1.21e{-2}$	$7.50e{-2}$
Linear	RealNVP	$6.19e3$	$1.54e2$	$3.54e11$
Linear	LegendreNet	$4.14e{-3}$	$3.10e{-2}$	$5.70e{-2}$
Linear	GCF	$8.42e{-3}$	$3.40e{-2}$	–
Linear	ContactNet	$5.96e{-3}$	$3.32e{-3}$	$3.58e1$
Linear	ContactNet-ODE	$5.09e{-3}$	$5.42e{-3}$	$4.38e{-1}$
Tanh	MLP	$1.99e{-3}$	$1.18e{-2}$	$6.81e{-2}$
Tanh	RealNVP	$2.05e4$	$9.35e1$	$4.34e11$
Tanh	LegendreNet	$3.69e{-3}$	$2.79e{-2}$	$2.49e{-2}$
Tanh	GCF	$8.19e{-3}$	$3.61e{-2}$	–
Tanh	ContactNet	$8.07e{-3}$	$2.24e{-3}$	$4.95e1$
Tanh	ContactNet-ODE	$4.93e{-3}$	$3.98e{-3}$	$2.91e{-1}$

The MLP has the lowest standard-test full-state error on both targets. Triangular ContactNet has the lowest action-to-position contrast error on both; ContactNet-ODE also improves on the MLP and GCF in this probe. The neural-ODE variant lowers mean full-state test error relative to the triangular variant, with overlapping seed ranges. Its wide-range means are substantially smaller, although individual seeds need not improve. Across all twelve neural-ODE runs, the maximum sampled contact defect is 1.33×10^{-15} and every test and wide rollout remains finite. The tanh target can be represented by one prolonged coupling, but training the three-pair architecture does not consistently recover that setting under the tested budgets.

F.4 Architecture and optimization ablations

The ablations use the quadratic-action 1D front dataset unless otherwise stated. They are exploratory single-seed studies, separate from the three-seed comparison in Appendix G. We report only contrasts whose settings and outcomes are recoverable from saved run arguments and metrics.

F.4.1 Free propagation and finite-depth factorization bias

For the conservative front benchmark, $H = p^2/2$ and the reference map over an interval h is

$$(x, u, p) \longmapsto (x + hp,\ u + \tfrac{1}{2}hp^2,\ p). \tag{32}$$

The GCF baseline implements the Hamiltonian class $H_{\mathrm{GCF}} = M(p)p^2/2 + V(x) + F(x)u$, with each term advanced by a tractable contact substep (Testa et al., 2025). For the ideal choice $M(p) = h/\tau$ and $V = F = 0$, one kinetic substep of duration τ gives Equation (32) exactly. The implemented random Fourier feature readout need only approximate this constant on the relevant momentum range. The observed fit therefore reflects both the factorization and the fixed-feature parameterization.

ContactNet instead composes strict lifted Hénon maps with prolonged base maps and, in the reported front configuration, optional vertical action blocks. Its strict factors can approximate the free map by Theorem 4.1, while the extra factors can be set to identity. A finite fit must nevertheless coordinate the strict maps and suppress unnecessary action-dependent factors. This is a structural interpretation of why GCF is favored by the free target; finite-depth approximation and optimization effects remain jointly present. In particular, replacing the implemented triangular (x, u) coupling with an x-fixed action coupling removes the momentum-dependent forward prolongation denominator $X_x + pX_u$: for $X = x$ it equals one. It does not add GCF's dedicated kinetic substep. The inverse of the x-fixed action coupling still divides by its positive action scale.

Table 11 probes this distinction on the conservative free-front data. All five arms use the same regenerated 40/8/15 profile split, three seeds, 120 Adam objective-and-gradient updates, validation-only one-step checkpoint selection, and 12-step test rollout and reference-jet inversion. The original x, u coupling and x-fixed width-32 coupling both retain the vertical ϕ block; this fixed-width pair changes the base coupling while also changing parameter count. The width-41 x-fixed variant approximately matches the original parameter count. Removing ϕ at width 50 yields an x-fixed action-affine variant with a similar count, closer to GCF's linear-in-u factor bias. GCF and the original ContactNet were rerun as controls on the same regenerated data; the table compares only these reruns.

Table 11: Free-front denominator ablation: three-seed means at 120 Adam updates. State MSE averages (x, u, p). Reference inverse applies twelve steps of the frozen inverse to exact reference terminal jets. All 15 jobs completed with finite coordinates.

Model	Params	One step	Step 12	Reference inverse
GCF	2,736	$7.30e{-5}$	$8.50e{-3}$	$1.11e{-2}$
ContactNet, $x, u + \phi$, width 32	2,739	$3.43e{-3}$	$5.03e{-1}$	$2.19e{-1}$
ContactNet, $x + \phi$, width 32	2,157	$3.05e{-3}$	$4.32e{-1}$	$3.74e{-1}$
ContactNet, $x + \phi$, width 41	2,751	$2.82e{-3}$	$3.97e{-1}$	$3.34e{-1}$
ContactNet, x, width 50	2,742	$2.76e{-3}$	$3.92e{-1}$	$3.24e{-1}$

Removing x, u mixing modestly reduces forward error at this budget, including in the approximately parameter-matched arms, but worsens reference inversion. Because the substitution changes the map class, parameterization, and denominator simultaneously, this experiment does not identify which change causes the difference. GCF remains substantially more accurate on free flow; isolating the effects of its Fourier features and optimizer dynamics would require separate controls.

An independent controlled sweep over $H = p^2/2 + x^2/2 + \gamma_k u^k/k$, $k = 1, 2, 3, 4$, provides a broader finite-capacity check. It uses three seeds, depths 1/2/4/8, two capacity views, and 400 Adam updates for each of 192 attempted fits. In the approximately fixed-total-parameter view at depth four, GCF has lower median one-step error for $k = 1, 3$, while ContactNet has lower median error for $k = 2, 4$. All attempts completed, and the ranking also varies with depth. The sweep establishes variation across the tested powers and depths; bracket-depth and approximation-rate laws would

require a separate analysis. The complete protocol, seed distributions, and cost views are summarized in Appendix F.3.

The theorem comparison is separate. Theorem 4.1 proves C^r density of variable-depth LEGENDRENET for smooth strict contact maps. Theorem 4.2 proves general contact-map approximation for sufficiently dense strict and base-map families, and Corollary 4.3 separates the proved neural-flow instantiation from the criterion for other base families. The available theory therefore gives a density guarantee for our neural-flow instantiation, while the implemented GCF random-feature splitting family comes without a corresponding density theorem (Testa et al., 2025). Neither result predicts which finite architecture, block vocabulary, or training budget best fits a given Hamiltonian. Adding a kinetic shear to our strict vocabulary, or a different action-dependent generator to another contact architecture, would change that finite-size comparison without changing the logical scope of these results.

F.4.2 BASE COUPLING AND LEGENDRE-ATOM SHARING

Table 12 varies two design choices in a six-run grid. Each model has five stages, four Legendre atoms per strict block, width 96, a residual vertical action block, and identity initialization. All use seed zero, 300 Adam epochs, one-step training, and the same 40 training and 15 held-out fronts. The base coupling is either an x-fixed action shear, a triangular x, u transformation, or a residual base-map variant. For the residual map, invertibility requires a Lipschitz condition and its inverse is not algebraic. The second axis ties or separates the four Legendre-atom parameter sets *within* each strict block. It does not tie parameters across network stages. The scale parameter follows the implementation for each base family: $k = 0.5$ for the first two and $k = 0.2$ for the residual base map.

Table 12: Single-seed component grid on quadratic-action fronts. State and Legendrian columns are held-out 12-step rollout metrics; final drift is state MSE at the last time. One-step state is evaluated on pooled *training* transitions. The six runs have the same data and nominal Adam schedule.

Base map	Legendre atoms	One step	Rollout state	Legendrian	Final drift
x-fixed action shear	Tied	$5.10e{-4}$	$5.25e{-3}$	$1.61e{-5}$	$1.83e{-2}$
x-fixed action shear	Independent	$4.21e{-4}$	$3.58e{-3}$	$1.61e{-5}$	$1.17e{-2}$
x, u triangular	Tied	$3.81e{-4}$	$3.60e{-3}$	$1.63e{-5}$	$1.24e{-2}$
x, u triangular	Independent	$1.04e{-4}$	$2.53e{-3}$	$1.49e{-5}$	$7.21e{-3}$
Residual base	Tied	$4.95e{-4}$	$3.92e{-3}$	$1.65e{-5}$	$1.32e{-2}$
Residual base	Independent	$2.08e{-4}$	$4.31e{-3}$	$1.48e{-5}$	$1.57e{-2}$

Independent atoms reduce the one-step error of the triangular x, u variant from 3.81×10^{-4} to 1.04×10^{-4}, a larger gain than for the x-fixed shear. The best 12-step state score in this grid is also the independent-atom triangular variant, 2.53×10^{-3}. These comparisons identify an interaction at this seed and training budget; persistence across seeds and optimization schedules was not tested. The Legendrian residual changes little across the six contact variants, as expected from their structure-preserving block formulas.

F.4.3 OPTIMIZATION BUDGET AND DEPTH PROBES

Table 13 follows the independent-atom, triangular CONTACTNET from the previous grid through longer optimization. The first three rows use the same seed and architecture. The Adam stage has 300 epochs, and the two subsequent rows run 300 or 2,000 *L-BFGS steps*. The MLP row is an Adam-only reference and has a different architecture and optimization budget. Each run keeps the logged checkpoint with the lowest held-out rollout error. For the 2,000-step run this is L-BFGS step 1,700, while every other run in Tables 12 and 13 keeps its final checkpoint.

The longer contact-model optimization lowers both one-step and 12-step error while leaving the sampled Legendrian residual near its discretization floor. The single-seed contact fit had not converged at the Adam checkpoint. The MLP row uses a different optimization budget. A separate depth probe uses a rational action block and tied Legendre atoms at width 96. Increasing from five to eight stages changes one-step state error from 3.83×10^{-4} to 3.73×10^{-4} and rollout state error from 3.97×10^{-3} to 3.76×10^{-3}. These two single-seed points provide a local depth contrast; an accuracy-versus-

Table 13: Single-seed optimization series on the quadratic dataset. One-step state is evaluated on pooled training transitions; rollout, drift, and Legendrian scores use held-out fronts. The ContactNet architecture is fixed across its three rows, and the 2,000-step row reports the checkpoint kept at L-BFGS step 1,700. The MLP is an unmatched Adam-only reference.

Training	One-step state	Rollout state	Final drift	Legendrian
CONTACTNET, Adam 300	$1.04e{-4}$	$2.53e{-3}$	$7.21e{-3}$	$1.49e{-5}$
CONTACTNET, Adam 300 + L-BFGS 300	$2.93e{-5}$	$5.54e{-4}$	$1.48e{-3}$	$1.44e{-5}$
CONTACTNET, Adam 300 + L-BFGS 2,000	$2.12e{-6}$	$2.02e{-5}$	$4.79e{-5}$	$1.37e{-5}$
MLP, Adam 300	$5.72e{-6}$	$9.60e{-5}$	$2.86e{-4}$	$7.29e{-4}$

depth curve would require additional depths and seeds. Both depth-probe runs also keep their final checkpoint.

Table 12 measures forward and sampled Legendrian errors under its recorded settings; blockwise effects on inversion and regularity require a separate ablation. The independent sparse-front study in Appendix G.1 addresses training-set size under a separate protocol.

G HAMILTON–JACOBI AND LEGENDRIAN-FRONT BENCHMARKS

G.1 SPARSE-FRONT DATA AND BUDGET SENSITIVITY

The target is $H = p^2/2 + x^2/2 + u^2/4$. A fixed dataset contains 40 candidate training, eight validation, and 15 test profiles, with 64 ordered rays and 13 times up to 1.2. All test fronts are folded at terminal time. The nested 2/4/8-front sets are prefixes of the training set. Reference reintegration with 64 instead of 16 RK4 substeps changes one tested trajectory by at most 1.40×10^{-10}. Every fit uses 600 Adam updates. The 300-step L-BFGS recipe includes MLP, RealNVP, GCF, and CONTACTNET; independent 1,000-step fits include MLP, RealNVP, and CONTACTNET. All models have approximately 2.7k parameters. Checkpoint selection uses only validation rollout error. Seeds are 0, 1, and 2. No 16-front experiment was run.

Table 14: Sparse-front terminal state MSE: arithmetic mean $\pm$ sample SD over three seeds. GCF failures are counted explicitly; finite-only means are not substituted. A dash denotes an unrun setting.

L-BFGS	Fronts	MLP	RealNVP	GCF	CONTACTNET
300	2	$3.55e{-2} \pm 6.26e{-3}$	$2.40e{-2} \pm 4.63e{-3}$	1/3 nonfinite	$5.11e{-3} \pm 1.37e{-3}$
300	4	$2.32e{-2} \pm 5.54e{-3}$	$1.39e{-2} \pm 2.75e{-3}$	2/3 nonfinite	$4.81e{-2} \pm 7.78e{-2}$
300	8	$1.65e{-2} \pm 6.33e{-3}$	$7.65e{-3} \pm 3.03e{-3}$	$9.17e1 \pm 1.55e2$	$4.01e{-3} \pm 1.98e{-3}$
300	40	$6.24e{-4} \pm 3.46e{-4}$	$7.96e{-4} \pm 1.52e{-4}$	$2.61e{-2} \pm 9.80e{-4}$	$8.47e{-4} \pm 3.09e{-4}$
1000	2	$2.18e{-2} \pm 1.55e{-3}$	$1.42e{-2} \pm 2.73e{-3}$	–	$9.11e{-3} \pm 8.61e{-3}$
1000	4	$9.15e{-3} \pm 1.72e{-3}$	$3.48e{-3} \pm 1.33e{-3}$	–	$2.35e{-3} \pm 9.23e{-4}$
1000	8	$5.15e{-3} \pm 8.68e{-4}$	$2.82e{-3} \pm 1.08e{-3}$	–	$3.23e{-3} \pm 2.64e{-3}$
1000	40	$1.10e{-4} \pm 3.15e{-5}$	$1.64e{-4} \pm 2.49e{-5}$	–	$1.65e{-4} \pm 9.73e{-5}$

All 84 fitting jobs completed, but the independently implemented GCF baseline showed pronounced seed-dependent blow-up at low data counts under the 300-step recipe: three selected test rollouts were nonfinite, and the finite errors varied by orders of magnitude. These outcomes diagnose the stated training recipe rather than asymptotic expressivity. The shorter-budget CONTACTNET result at four fronts is likewise dominated by one seed; its independent 1,000-step fit is finite and substantially smaller. The main-text figure therefore emphasizes the matched 1,000-step MLP, RealNVP, and CONTACTNET comparison while retaining the 300-step optimizer-budget view.

At eight fronts, CONTACTNET wins two of three paired seed comparisons against RealNVP while having a worse arithmetic mean; one poorer fit reverses the aggregate ranking.

Exact-terminal-jet inverse scores follow the same mixed ranking as the forward errors and show no robust advantage over RealNVP (Table 15). Numerically lifted terminal fronts introduce momentum RMSE of about 0.024 and are a distinct test addressed by the complete-front recovery experiment below.

Table 15: Sparse-front reference-terminal-jet inverse recovery. Entries are three-seed mean initial-state MSE where all inverses are finite. Counts replace means for incomplete GCF settings. The MLP has no architectural inverse. These exact-jet scores exclude momentum reconstruction from projected fronts.

L-BFGS	Fronts	RealNVP	GCF	CONTACTNET
300	2	$4.94e{-2}$	2/3 finite	$5.76e{-3}$
300	4	$2.67e{-2}$	1/3 finite	$5.54e{-2}$
300	8	$8.91e{-3}$	$2.20e{-1}$	$6.95e{-3}$
300	40	$6.79e{-4}$	$3.39e{-2}$	$6.98e{-4}$
1000	2	$1.99e{-2}$	–	$2.52e{-2}$
1000	4	$5.14e{-3}$	–	$4.67e{-3}$
1000	8	$3.18e{-3}$	–	$6.90e{-3}$
1000	40	$1.30e{-4}$	–	$1.57e{-4}$

The sampled Legendrian MSE is the average of $(U_s - PX_s)^2$ over the predicted rollout. Its reference floor is 3.25×10^{-4}. For 2/4/8/40 fronts at 1,000 steps, CONTACTNET gives $3.03/3.49/3.42/3.32 \times 10^{-4}$; RealNVP gives $1.83e{-1}/4.77e{-2}/4.30e{-2}/4.97e{-3}$.

G.2 SEEDED ONE-DIMENSIONAL FRONT ROLLOUTS AND INVERSION

One-dimensional benchmark. We evaluate the fitted maps on fronts generated by

$$H(x, u, p) = \tfrac{1}{2}e^{\beta u}p^2 + \tfrac{1}{2}\omega^2 x^2 + \frac{\gamma}{k}u^k. \tag{33}$$

The three regimes are conservative free propagation $(\omega, \beta, \gamma) = (0, 0, 0)$, quadratic action dependence $(\omega, \beta, \gamma, k) = (1, 0, 0.5, 2)$, and intensity dependence $(\omega, \beta, \gamma) = (1, 0.5, 0)$. The conservative case has no quadratic potential, so the regimes do not vary in action dependence alone. Each uses 40 training and 15 held-out fronts, 64 ordered points per front, and 13 times from 0 to 1.2. Models fit adjacent lifted states and run 12 autoregressive steps from each held-out initial front. These are long autoregressive rollouts from one-step training, not tests beyond the training time interval.

The five-model comparison has three initialization seeds. LEGENDRENET, GCF, and CONTACTNET use the same one-step data and nominal schedule of 150 Adam epochs followed by 300 L-BFGS steps. CONTACTNET also uses its nonsingularity penalties. The reported RealNVP runs have 500 L-BFGS steps, and the MLP has 42.8k parameters versus about 13.5k for each invertible model. We therefore do not treat all five as optimization matched. Checkpoints were selected by held-out rollout error without an independent validation set, so selection and reporting use the same profiles; Table 17 gives the final-checkpoint comparison. We measure full-state rollout MSE and the mean squared finite-difference Legendrian residual $u_s - px_s$.

Table 16: One-dimensional held-out rollout, three-seed means. Each regime shows state MSE and Legendrian residual MSE (Leg.). Bold marks the lowest unrounded mean in each column. The MLP has an enlarged parameter budget and RealNVP has a longer L-BFGS schedule.

	Conservative		Quadratic		Intensity	
Model	State	Leg.	State	Leg.	State	Leg.
MLP	$2.7e{-5}$	$3.02e{-4}$	$1.1e{-4}$	$8.81e{-4}$	$7.7e{-5}$	$7.20e{-4}$
RealNVP	$1.4e{-5}$	$1.85e{-4}$	$\mathbf{5.5e{-5}}$	$5.88e{-4}$	$\mathbf{6.8e{-5}}$	$3.65e{-4}$
LEGENDRENET	$8.8e{-5}$	$2.71e{-5}$	$2.7e{-2}$	$1.56e{-5}$	$3.5e{-2}$	$\mathbf{1.47e{-5}}$
GCF	$\mathbf{4.5e{-8}}$	$\mathbf{2.66e{-5}}$	$6.0e{-3}$	$2.44e{-5}$	$1.7e{-4}$	$1.54e{-5}$
CONTACTNET	$2.0e{-5}$	$2.70e{-5}$	$1.1e{-4}$	$\mathbf{1.39e{-5}}$	$8.0e{-4}$	$1.54e{-5}$

The forward winner varies with the target. GCF closely matches free propagation, while RealNVP has the lowest state MSE in the two action-dependent regimes under these reported budgets. CONTACTNET is close to the larger MLP on the quadratic system but less accurate on the intensity system. The strict LEGENDRENET has much larger state error when action feeds back into momentum. All three contact architectures have Legendrian residuals around 10^{-5}, while RealNVP and the MLP are around 10^{-4} to 10^{-3}. The residual is computed from sampled curves, so even exactly contact maps have a finite-difference floor.

Checkpoint selection. Table 17 measures the effect of selecting on the held-out fronts, using the logged training histories. Keeping the last logged checkpoint instead preserves the order of the five models in every column of Table 16. Each three-seed mean state or Legendrian MSE changes by about 5% or less, except intensity LEGENDRENET state MSE, which remains last. At the matched 300-step L-BFGS budget the MLP has the lowest intensity state MSE, 7.74×10^{-5} against RealNVP's 9.75×10^{-5}, so RealNVP's lead in that regime depends on its longer schedule. This check covers forward metrics only. Because only the selected checkpoint weights were archived, the inverse diagnostics in Table 18 use those weights; final-checkpoint inverse diagnostics cannot be reconstructed for runs that selected an earlier checkpoint. Selection used held-out forward rollout error, not an inverse diagnostic. Among the 36 invertible-model runs, 21 selected the final checkpoint. All three quadratic GCF runs selected the final checkpoint, whereas the quadratic CONTACTNET runs selected epochs 425, 425, and 375 of 450.

Table 17: Checkpoint-selection sensitivity for Table 16, three-seed means. Selected is the reported checkpoint, with the lowest logged held-out rollout state MSE. Final is the last logged checkpoint. Matched follows 150 Adam epochs and 300 L-BFGS steps and differs from Final only for RealNVP, whose reported runs take 500 steps.

		State MSE			Legendrian MSE	
Regime	Model	Selected	Final	Matched	Selected	Final
Conservative	MLP	$2.67e{-5}$	$2.67e{-5}$	$2.67e{-5}$	$3.02e{-4}$	$3.02e{-4}$
	RealNVP	$1.43e{-5}$	$1.43e{-5}$	$1.93e{-5}$	$1.85e{-4}$	$1.85e{-4}$
	LEGENDRENET	$8.80e{-5}$	$8.80e{-5}$	$8.80e{-5}$	$2.71e{-5}$	$2.71e{-5}$
	GCF	$4.49e{-8}$	$4.69e{-8}$	$4.69e{-8}$	$2.66e{-5}$	$2.65e{-5}$
	CONTACTNET	$2.02e{-5}$	$2.06e{-5}$	$2.06e{-5}$	$2.70e{-5}$	$2.70e{-5}$
Quadratic	MLP	$1.05e{-4}$	$1.05e{-4}$	$1.05e{-4}$	$8.81e{-4}$	$8.81e{-4}$
	RealNVP	$5.47e{-5}$	$5.54e{-5}$	$6.55e{-5}$	$5.88e{-4}$	$5.79e{-4}$
	LEGENDRENET	$2.71e{-2}$	$2.72e{-2}$	$2.72e{-2}$	$1.56e{-5}$	$1.54e{-5}$
	GCF	$6.00e{-3}$	$6.00e{-3}$	$6.00e{-3}$	$2.44e{-5}$	$2.44e{-5}$
	CONTACTNET	$1.09e{-4}$	$1.14e{-4}$	$1.14e{-4}$	$1.39e{-5}$	$1.39e{-5}$
Intensity	MLP	$7.73e{-5}$	$7.74e{-5}$	$7.74e{-5}$	$7.20e{-4}$	$7.06e{-4}$
	RealNVP	$6.85e{-5}$	$6.86e{-5}$	$9.75e{-5}$	$3.65e{-4}$	$3.65e{-4}$
	LEGENDRENET	$3.50e{-2}$	$4.34e{-2}$	$4.34e{-2}$	$1.47e{-5}$	$1.41e{-5}$
	GCF	$1.73e{-4}$	$1.73e{-4}$	$1.73e{-4}$	$1.54e{-5}$	$1.54e{-5}$
	CONTACTNET	$8.01e{-4}$	$8.02e{-4}$	$8.02e{-4}$	$1.54e{-5}$	$1.52e{-5}$

Table 18 gives the aggregate inverse results for the same fitted model groups. All models receive the same numerically lifted terminal front in a given regime, so the terminal momentum-lift MSE does not vary with training seed: it is 4.62×10^{-3} for conservative propagation, 2.46×10^{-4} for quadratic action dependence, and 2.83×10^{-4} for intensity dependence. The MLP has no architectural inverse and therefore has no row. Small self-cycle error confirms that the implemented inverse undoes its own fitted map on tested states. Recovery from a reference terminal front is the separate task measured by the trajectory errors.

Table 18: One-dimensional inverse recovery: three-seed mean $\pm$ population standard deviation for the backward trajectory from a numerically lifted reference terminal front. Cycle MSE measures a model forward–inverse self-cycle. Recovered Leg. is the mean squared finite-difference residual of the recovered initial jet after 12 inverse steps, in float64; dashes denote regimes not evaluated for this diagnostic.

Regime	Model	Recovery MSE	Cycle MSE	Recovered Leg.
Cons.	RealNVP	$7.28e{-3} \pm 2.48e{-6}$	$1.43e{-13} \pm 1.25e{-14}$	–
Cons.	LEGENDRENET	$7.30e{-3} \pm 1.17e{-4}$	$2.88e{-12} \pm 1.22e{-12}$	–
Cons.	GCF	$7.28e{-3} \pm 1.56e{-6}$	$5.38e{-13} \pm 1.93e{-14}$	–
Cons.	CONTACTNET	$7.40e{-3} \pm 2.15e{-5}$	$1.64e{-12} \pm 2.11e{-13}$	–
Quadratic	RealNVP	$2.56e{-4} \pm 1.48e{-5}$	$2.11e{-13} \pm 2.46e{-14}$	$9.25e{-4} \pm 1.40e{-4}$
Quadratic	LEGENDRENET	$3.06e{-2} \pm 2.38e{-4}$	$4.88e{-12} \pm 4.72e{-12}$	$3.09e{-5} \pm 2.79e{-7}$
Quadratic	GCF	$5.99e{-3} \pm 7.44e{-4}$	$7.88e{-13} \pm 1.03e{-14}$	$2.46e{-5} \pm 1.32e{-6}$
Quadratic	CONTACTNET	$2.75e{-4} \pm 3.33e{-5}$	$1.54e{-12} \pm 3.05e{-14}$	$2.35e{-5} \pm 5.00e{-7}$
Intensity	RealNVP	$4.79e{-4} \pm 3.52e{-5}$	$3.13e{-13} \pm 6.08e{-16}$	–
Intensity	LEGENDRENET	$2.97e{-2} \pm 1.58e{-3}$	$9.35e{-13} \pm 3.11e{-13}$	–
Intensity	GCF	$6.98e{-4} \pm 4.30e{-5}$	$8.24e{-13} \pm 4.41e{-14}$	–
Intensity	CONTACTNET	$9.87e{-4} \pm 2.67e{-4}$	$2.79e{-12} \pm 6.33e{-13}$	–

For the recovered-jet diagnostic, we promote archived weights to float64 while preserving GCF's original random-feature basis. Each checkpoint reproduces its stored forward score before evaluation. The residual is $\mathrm{mean}_s(D_s\widehat{u}_0 - \widehat{p}_0 D_s\widehat{x}_0)^2$, using the physical ray grid and the same one-sided endpoint convention as the forward metric. The initial reference residual is 1.79×10^{-5}; CONTACTNET's recovered residual is 2.35×10^{-5} versus RealNVP's 9.25×10^{-4}. These are three-seed means after averaging 15 fronts per seed. The contact inverse preserves exact Legendrian input curves; numerical lifting and finite differencing explain the remaining residual.

RealNVP and CONTACTNET have quadratic backward-trajectory errors near the terminal-lift error scale. GCF's quadratic recovery error is about 22 times CONTACTNET's, but their forward fit differs by about 55 times under the reported budgets. In intensity GCF recovers more accurately than CONTACTNET, and all four invertible models have similar conservative recovery error. The strict LEGENDRENET has large recovery error when the reference dynamics depend on action. These results combine fitted-map error with inversion from a lifted reference front. The perturbation experiment below separately probes local amplification.

G.3 INVERSE AND REGULARITY DIAGNOSTICS

An auxiliary short-budget check used fresh conservative and quadratic-front datasets with disjoint 40/8/15 profile splits. Four invertible architectures, three seeds, and 15 test fronts produced 360 finite backcasts. The maximum model self-cycle error was 5.7×10^{-15} despite appreciable error when backcasting reference terminal states, confirming that algebraic reversal and physical recovery measure different properties. For signed tangent and transverse terminal perturbations of magnitude 10^{-3}, the median measured backcast amplification was 1.066, with range 0.516–4.752. These local finite-perturbation measurements quantify the evaluated neighborhoods; a global condition-number bound would require control over the full domain, which exact contactness alone does not supply.

G.4 COMPLETE-FRONT INVERSE RECOVERY IN ONE DIMENSION

G.4.1 PER-FRONT RECOVERY FROM COMPLETE TERMINAL FRONTS

The higher-budget CONTACTNET case study in Section 7 uses one frozen model and all fifteen held-out fronts of the quadratic-action dataset. The terminal observation for each case is the full ordered (X_T, U_T) curve at $T = 1.2$. Momentum is estimated with the numerical lift in Appendix H, and the resulting jet is passed through twelve exact architectural inverse steps. No parameters are fitted during this recovery.

Table 19 reports every held-out case. A fold step is the first sampled time index at which any adjacent pair of ordered rays has decreasing x; a dash means no such pair occurs through T. With time spacing 0.1, the recorded folded cases first meet this criterion between indices 3 and 11. This is a discrete fold-detection rule, not a continuous-time estimate of the first caustic. The branch count splits the terminal curve at changes in the sign of adjacent x differences.

Twelve fronts are folded at the terminal time. For those cases, terminal lift MSE has median 5.70×10^{-5}, and backward-trajectory state MSE has median 6.29×10^{-5}. The three unfolded fronts have a backward-trajectory median of 3.04×10^{-5}. The folded group includes a clear difficult case: front 12 has terminal lift MSE 2.96×10^{-3} and backward-trajectory MSE 2.15×10^{-3}. Its recovered initial profile visibly departs from the reference near the right endpoint in Figure 7. The discrepancy combines error from the terminal lift, the learned map, and their interaction during inversion; the separate metrics report its stages without attributing a single cause.

The self-cycle MSEs for folded cases range from 1.02×10^{-12} to 4.14×10^{-12}, including 2.75×10^{-12} for front 12. These values check algebraic reversal of each model's own forward trajectory. The much larger reference-trajectory errors in the same rows measure a different task. The first detected fold step and the terminal branch count describe the geometry coarsely. Across these cases, recovery error has no monotone association with either quantity.

Figure 6 shows the complete lifting and inverse-recovery procedure.

Table 19: All fifteen held-out fronts for one higher-budget CONTACTNET. The terminal lift column is MSE of $\widehat{P}_T$ against reference P_T over 64 ray samples. Backward state and action columns average over all thirteen states of the recovered trajectory; state is the sum of the x, u, and p coordinate MSEs. Cycle MSE applies the same fitted model forward and backward from the reference initial jet. Fold step and terminal branch count use the sampled x curve. Rows with a dash are not folded at the terminal time.

Front	Fold step	Branches at T	Terminal lift p MSE	Backward state MSE	Backward u MSE	Self-cycle MSE
0	9	4	$7.58e{-5}$	$6.36e{-5}$	$4.16e{-6}$	$1.02e{-12}$
1	–	1	$3.46e{-5}$	$2.86e{-5}$	$3.62e{-6}$	$8.20e{-13}$
2	–	1	$3.64e{-5}$	$2.85e{-5}$	$3.25e{-6}$	$9.50e{-13}$
3	10	2	$5.06e{-5}$	$5.88e{-5}$	$1.93e{-5}$	$3.06e{-12}$
4	11	3	$5.77e{-5}$	$5.07e{-5}$	$5.53e{-6}$	$1.18e{-12}$
5	3	3	$6.29e{-5}$	$8.47e{-5}$	$2.56e{-5}$	$4.14e{-12}$
6	6	2	$4.63e{-5}$	$6.21e{-5}$	$1.42e{-5}$	$2.98e{-12}$
7	9	3	$3.76e{-5}$	$4.94e{-5}$	$9.08e{-6}$	$2.82e{-12}$
8	–	1	$3.64e{-5}$	$3.04e{-5}$	$3.66e{-6}$	$9.02e{-13}$
9	5	3	$4.90e{-5}$	$4.93e{-5}$	$5.83e{-6}$	$1.35e{-12}$
10	9	2	$6.33e{-5}$	$9.62e{-5}$	$2.97e{-5}$	$2.68e{-12}$
11	4	3	$5.00e{-5}$	$7.48e{-5}$	$2.02e{-5}$	$2.73e{-12}$
12	5	3	$2.96e{-3}$	$2.15e{-3}$	$6.37e{-4}$	$2.75e{-12}$
13	6	3	$5.63e{-5}$	$6.95e{-5}$	$1.07e{-5}$	$2.34e{-12}$
14	11	3	$6.67e{-5}$	$6.04e{-5}$	$6.69e{-6}$	$1.18e{-12}$

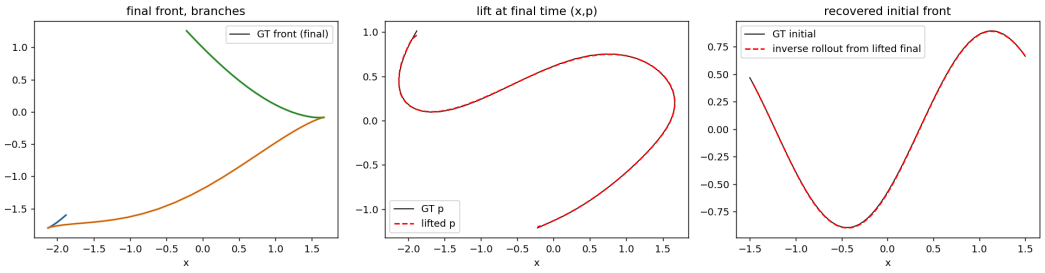

Figure 6: One-dimensional recovery from one held-out folded front. The terminal (x, u) curve is lifted to estimate p, then the frozen CONTACTNET inverse recovers the initial front. Black curves show the reference where overlaid.

G.4.2 OBSERVATION NOISE BEFORE LIFTING

The higher-budget frozen model is evaluated on the same 12 folded fronts. Independently perturb each terminal x and u sample by $N(0, \sigma^2)$ noise, in the coordinate units of the benchmark, at $\sigma = 10^{-4}, 10^{-3}, 10^{-2}$, with three draws per front. The clean row repeats the unperturbed input. The raw lift uses the existing threshold $0.25\,\mathrm{median}\,|D_s x|$; the smoothed lift first applies a Savitzky–Golay filter to both coordinates (window nine rays, polynomial degree three, polynomial endpoint extension), then uses the same ratio and masking rule. This rule is fixed across all noise levels. Paired raw and smoothed runs use the same noise arrays; levels scale the same standard-normal draws, seeded by $23092026 + 100d + i$ for draw $d = 0, 1, 2$ and front index i. Table 20 reports medians over the 36 front/draw cases, not independent training-seed uncertainty. No model is retrained.

Smoothing reduces the median recovery error at each nonzero noise level, but the largest level remains far from the clean accuracy. This is consistent with the derivative-error amplification in Proposition H.1, not a uniform statistical stability claim.

G.4.3 OBSERVATION ASSUMPTIONS

These runs use one 64-point ray grid and the fixed near-cusp threshold $|x_s| \le 0.25\,\mathrm{median}\,|x_s|$. The noise sweep above varies observation noise with a fixed smoothing rule; ray resolution and cusp threshold are not varied. The sweep therefore measures sensitivity for this discretization. Terminal

Table 20: Noise before the terminal lift: median momentum-lift MSE and backward-trajectory state MSE. Recovery sums the three coordinate MSEs over all 13 backward states; finite counts include every case. At zero noise the three draws coincide.

σ	Lift	Momentum MSE	Recovery MSE	Finite
0	Raw	$5.70e{-5}$	$6.29e{-5}$	36/36
0	Smoothed	$5.62e{-5}$	$6.94e{-5}$	36/36
$1.00e{-4}$	Raw	$1.40e{-4}$	$1.33e{-4}$	36/36
$1.00e{-4}$	Smoothed	$7.22e{-5}$	$8.49e{-5}$	36/36
$1.00e{-3}$	Raw	$4.69e{-3}$	$3.54e{-3}$	36/36
$1.00e{-3}$	Smoothed	$1.19e{-3}$	$1.42e{-3}$	36/36
$1.00e{-2}$	Raw	$3.28e{-1}$	$2.79e{-1}$	36/36
$1.00e{-2}$	Smoothed	$7.02e{-2}$	$5.92e{-2}$	36/36

momentum error and backward-trajectory error are reported separately to distinguish the two stages of recovery.

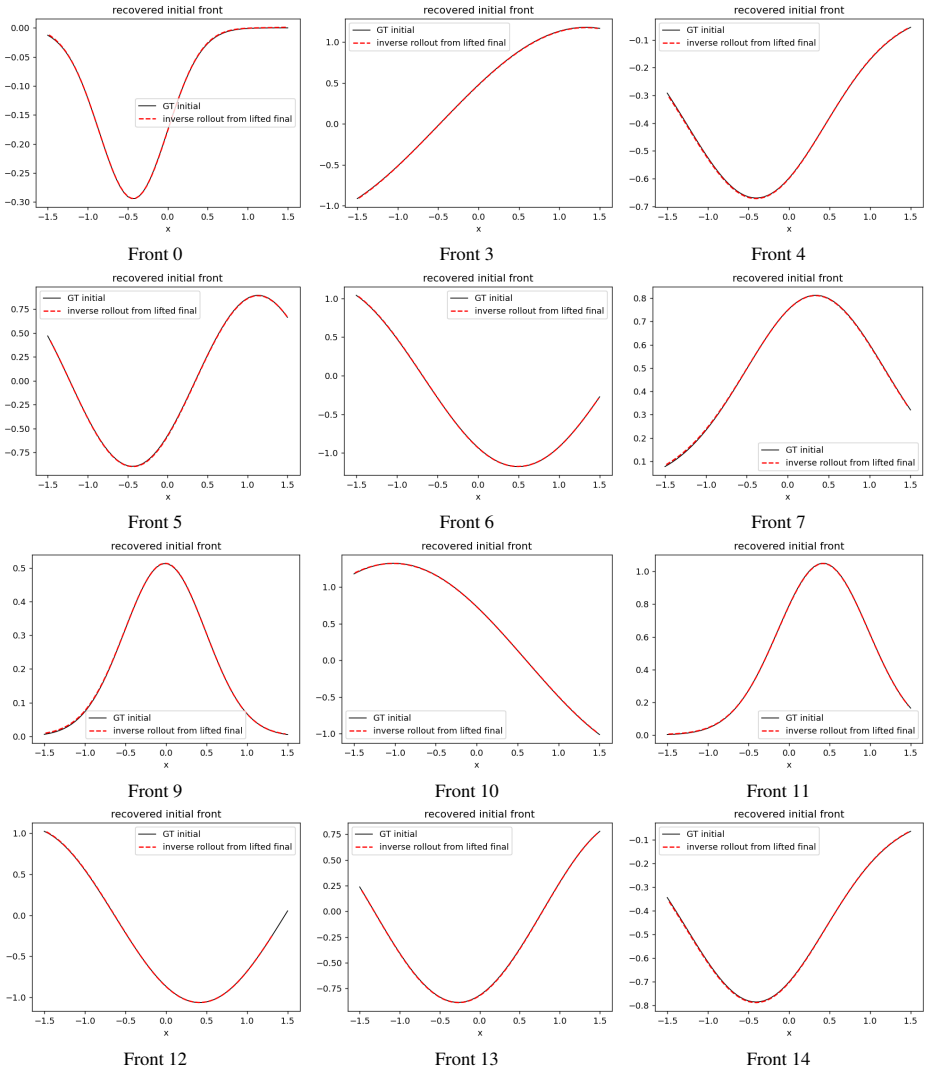

Figure 7: Recovered initial profiles for all twelve terminally folded held-out fronts under one higher-budget CONTACTNET. Black denotes the reference and red dashes the result of lifting the ordered terminal front and applying the frozen inverse. Front 12 shows the visible failure case. See Table 19 for the corresponding errors.

G.5 TWO-DIMENSIONAL OPTICS EXPERIMENTS

G.5.1 FLOW, ARCHITECTURE, AND TRAINING

The two-dimensional benchmark uses $J^1(\mathbb{R}^2)$ with $x, p \in \mathbb{R}^2$, $u \in \mathbb{R}$, and

$$H(x, u, p) = \tfrac{1}{2}\|p\|^2 + \tfrac{1}{4}u^2. \tag{34}$$

The configuration has zero spatial lens strength; focusing comes from the initial fronts. Since $H_u = u/2$, the momentum equation contains $-(u/2)p$ and the (x, p) dynamics depend on action. Reference characteristics use sixteen RK4 substeps between each pair of thirteen equally spaced times on $[0, 1.2]$.

Each initial jet is $(s, f_0(s), \nabla f_0(s))$ on a 20×20 Lagrangian grid $s \in [-1, 1]^2$. Thirty sampled profiles train one step map. The test set has ten further sampled profiles plus a designated Gaussian profile, for eleven held-out profiles. The sampled profiles mix astigmatic focusing fronts and Gaussian bumps. Their parameters and ordering are stored in the run metrics.

The contact model has ten stages of strict Legendre and prolonged base blocks. Each strict block has two Legendre layers with six independently parameterized Hénon-like atoms per layer. The base block combines an x-fixed action shear with a u-conditioned spatial map $X = a(u)x + b(u)$, $U = u$. Its momentum denominator is

$$d = a(u) + p \cdot \big(a'(u)x + b'(u)\big). \tag{35}$$

The fields use width 96. The unconstrained comparison is a width-96, three-layer residual MLP on all five jet coordinates. The strict comparison has ten symplectic layers of width 96 and is scored only on (x, p).

All three models fit adjacent lifted states with 120 AdamW epochs at learning rate 2×10^{-3}. The one-step loss includes a weight-three term matching the finite-difference mesh Jacobian $D_s X$. The contact model also receives a denominator hinge penalty with weight one and margin 0.2. It attempts 500 full-batch L-BFGS steps after AdamW, but its selected checkpoint is AdamW epoch 120. The MLP and strict checkpoints are epochs 100 and 60, respectively, without an L-BFGS phase. Checkpoint choice used the same eleven reported test profiles. The last logged checkpoints give state MSE 6.53×10^{-5} for CONTACTNET and 9.86×10^{-3} for the MLP, the same ratio of 151, and strict (x, p) MSE 1.54×10^{-1} against 1.43×10^{-1}. Seed-to-seed variability is not measured in this experiment.

G.5.2 PER-PROFILE ROLLOUTS

Table 21 summarizes the three architecture comparisons.

Table 21: Two-dimensional held-out rollout means across 11 profiles from one trained model per architecture. These are not seed aggregates.

Model	Front MSE	State MSE	Legendrian MSE
CONTACTNET	$3.01e{-5}$	$6.49e{-5}$	$1.32e{-3}$
MLP	$4.63e{-3}$	$9.81e{-3}$	$1.07e{-2}$
LEGENDRENET	(x, p) MSE $= 1.43e{-1}$ only		

Table 22 averages each metric over all thirteen times and 400 labeled rays. State MSE sums the x, u, and p coordinate-group MSEs. Front MSE sums the x and u terms. The Legendrian residual sums the squared finite-difference values $U_{s_i} - P \cdot X_{s_i}$ for $i = 1, 2$. It has a discretization floor even for an exactly contact map. The strict model's (x, p) score has a different output space and cannot be read as full-state MSE.

Table 22: All eleven held-out 2D profiles from one trained run per architecture. G denotes Gaussian and A astigmatic. Test 10 is the designated Gaussian appended to the ten sampled test profiles. C and MLP report rollout metrics; S reports only strict (x, p) MSE.

Test	Type	C state	C front	C Leg.	MLP state	MLP Leg.	S (x, p)
0	G	$2.92e{-4}$	$1.42e{-4}$	$6.31e{-3}$	$2.86e{-2}$	$3.66e{-2}$	$3.80e{-1}$
1	G	$7.25e{-6}$	$2.75e{-6}$	$2.14e{-3}$	$1.83e{-2}$	$2.01e{-2}$	$9.80e{-2}$
2	G	$3.57e{-6}$	$2.22e{-6}$	$1.29e{-4}$	$2.66e{-4}$	$7.10e{-4}$	$1.57e{-1}$
3	A	$2.79e{-5}$	$1.58e{-5}$	$1.28e{-4}$	$4.41e{-3}$	$3.07e{-3}$	$4.61e{-2}$
4	A	$4.24e{-5}$	$2.36e{-5}$	$2.13e{-4}$	$7.38e{-3}$	$1.42e{-2}$	$3.65e{-2}$
5	A	$1.34e{-5}$	$5.40e{-6}$	$9.78e{-5}$	$2.10e{-3}$	$1.45e{-3}$	$5.40e{-2}$
6	A	$3.15e{-5}$	$1.67e{-5}$	$1.91e{-4}$	$6.10e{-3}$	$5.38e{-3}$	$3.24e{-2}$
7	A	$1.64e{-5}$	$7.37e{-6}$	$1.02e{-4}$	$2.51e{-3}$	$2.05e{-3}$	$5.05e{-2}$
8	A	$1.13e{-5}$	$4.65e{-6}$	$1.64e{-4}$	$2.97e{-3}$	$5.75e{-3}$	$4.30e{-2}$
9	G	$6.43e{-5}$	$2.49e{-5}$	$2.69e{-4}$	$1.19e{-2}$	$2.90e{-3}$	$2.93e{-1}$
10	G	$2.04e{-4}$	$8.50e{-5}$	$4.79e{-3}$	$2.35e{-2}$	$2.60e{-2}$	$3.79e{-1}$
Mean		$6.49e{-5}$	$3.01e{-5}$	$1.32e{-3}$	$9.81e{-3}$	$1.07e{-2}$	$1.43e{-1}$

The contact model has lower state error than the MLP on every listed profile under these checkpoints. Its state MSE ranges from 3.57×10^{-6} on test 2 to 2.92×10^{-4} on test 0. Larger sampled Legendrian residuals on tests 0 and 10 also show dependence on front shape and resolution; this diagnostic is a finite-difference front residual rather than an algebraic contact-pullback defect. Representative projected meshes appear in Figures 4 and 8, and Table 22 reports the full-jet comparison.

G.5.3 Full-jet inversion and its limits

The inverse visualization selects held-out test 1, the earliest-folding test profile, whose projected x mesh first folds at about $t = 0.30$. It supplies the *reference terminal full jet* (X_T, U_T, P_T) at $T = 1.2$ and applies the frozen ContactNet inverse twelve times. The recovered initial (x, u) front has group-summed MSE 4.14×10^{-5}. In a separate model self-cycle, the maximum absolute coordinate difference after twelve forward and twelve inverse steps is 7.93×10^{-6}. Both values concern one profile and one trained map.

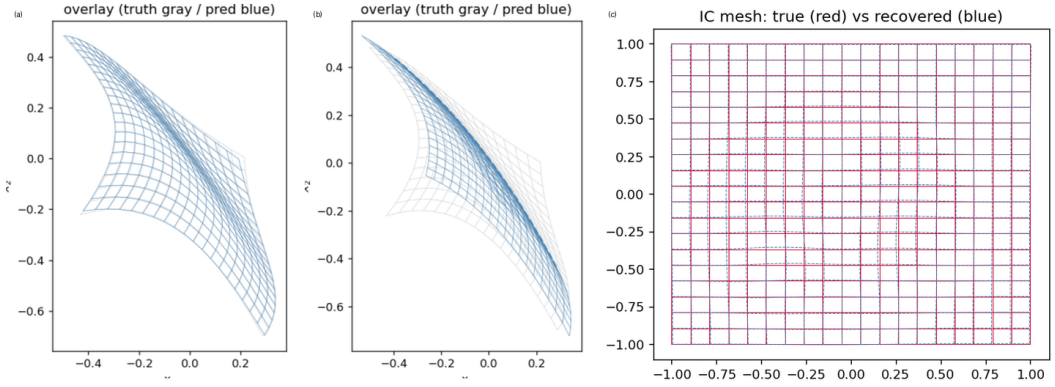

Figure 8: Two-dimensional optical caustics. Panels (a) and (b) compare the terminal ray mesh of held-out profile 5, the middle test index, predicted by ContactNet and the MLP, respectively, with the reference mesh in gray. Panel (c) compares true and recovered initial meshes for held-out test 1, the earliest-folding profile. The inversion is a single held-out case from one fitted model.

The terminal P_T assumption is essential here. At a projected caustic, $D_s X$ loses rank, so $D_s U = P D_s X$ does not alone determine P pointwise. This experiment evaluates momentum-supplied full-jet recovery; projected-only momentum recovery in two dimensions remains untested. The network block denominator (35) has a separate role: the algebraic inverse requires it to stay nonzero along the evaluated states. The training hinge discourages small $|d|$ on sampled transitions, but denominator minima, near-zero counts, and inverse-trajectory traces were not logged. Because rank loss in the *front projection* and a zero of a *network block* denominator are different events, this single-case result establishes inversion only along the evaluated full-jet trajectory; it does not certify global invertibility.

H Hamilton–Jacobi Front Lifting

H.1 Recovering the momentum of an ordered curve

Proposition 6.1 concerns the complete ordered front, with its curve parameter s retained. It does not assume that u is a single-valued function of x. At a fixed x, distinct front branches can therefore carry distinct values of both u and p.

Proof of Proposition 6.1. Since the curve is Legendrian, the pullback of $\alpha = du - p\,dx$ to its parameter domain vanishes. Hence

$$u_s(s) = p(s)x_s(s). \tag{36}$$

At every regular point of the projection to x, division by x_s gives the first formula in Equation (9). Let s_0 be a zero of x_s. Equation (36) also gives $u_s(s_0) = 0$. Differentiating once and evaluating at s_0 yields

$$u_{ss}(s_0) = p_s(s_0)x_s(s_0) + p(s_0)x_{ss}(s_0) = p(s_0)x_{ss}(s_0). \tag{37}$$

The assumed nonzero $x_{ss}(s_0)$ gives the second formula. Moreover, x_s has a simple zero at s_0, so it is nonzero in a punctured neighborhood of that point. On both sides, $u_s/x_s = p$, and continuity of p gives the same limit $p(s_0) = u_{ss}(s_0)/x_{ss}(s_0)$. Every zero of x_s is isolated. The regular points are consequently dense, and the recovered values there uniquely determine a continuous momentum at the zeros as well. □

The model curve

$$x(s) = s^2, \qquad u(s) = \tfrac{2}{3}s^3, \qquad p(s) = s \tag{38}$$

illustrates the continuation. It satisfies $u_s = p x_s$. For every $x > 0$ near the cusp, the two branches have $s = \pm\sqrt{x}$ and different values of u and p. Yet the ordered front determines $p = u_s/x_s = s$ away from $s = 0$, and $p(0) = u_{ss}(0)/x_{ss}(0) = 0$. The point of the ordering is to keep those two branches separate while following the curve through the cusp.

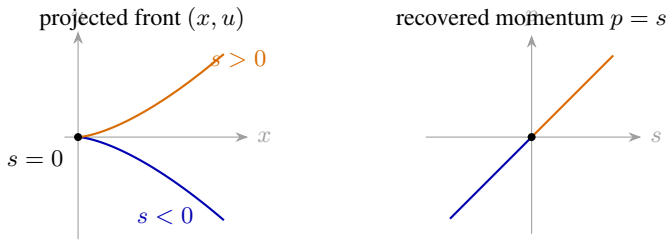

Figure 9: Local cusp model (38). The two colored branches of the (x, u) projection meet at a cusp. Their ratios u_s/x_s both tend to the same momentum, which remains smooth in the curve parameter. The sketch depicts a local model, not a benchmark trajectory.

H.2 Numerical lift used in the one-dimensional inverse task

The benchmark observes 64 samples of each terminal (X_T, U_T) front in ray order. For each coordinate, the lifting code uses `numpy.gradient` along the sample index to estimate its parameter

derivative. This gives centered differences at interior samples and one-sided differences at the endpoints. Because the same uniform index spacing is used for both derivatives, its scale cancels in their ratio. With d_i denoting the estimated derivative of X_T and $m = \mathrm{median}_i\,|d_i|$, the code computes

$$\widehat{P}_{T,i} = \frac{\widehat{(U_T)_{s_i}}}{d_i} \quad \text{where} \quad |d_i| > 0.25m. \tag{39}$$

It linearly interpolates $\widehat{P}_T$ in sample index over the remaining near-cusp samples, using the ratios on either side. If a masked region reaches an endpoint, the interpolation routine uses the nearest available ratio. A degenerate case with $m = 0$ instead uses the regularized ratio $\widehat{(U_T)_{s_i}} d_i/(d_i^2 + 10^{-12})$. The standard inverse runs keep the supplied ray order. An optional point-cloud ordering routine exists in the code, but its ability to recover the correct order is not part of the reported benchmark.

The interpolation is a numerical choice, not the second-derivative formula at a cusp. Its need is visible from the exact perturbation identity, at a regular point where $u_s = p x_s$:

$$\frac{u_s + \delta u_s}{x_s + \delta x_s} - p = \frac{\delta u_s - p\,\delta x_s}{x_s + \delta x_s}. \tag{40}$$

Derivative error is amplified when the denominator is small. Finite sampling, endpoint differences, observation noise, and interpolation bias across a wide masked neighborhood can all affect the terminal jet. Proposition H.1 quantifies these two contributions. In the inverse experiment, terminal lift error then enters every step of the frozen model inverse, along with error from the fitted map itself. The measured terminal momentum errors and backward-trajectory recovery errors are therefore reported separately in Section 7 and Table 18.

Proposition H.1 (Pointwise stability of the masked lift)**.** *At a regular point of an exact Legendrian curve, suppose derivative estimates satisfy $|\delta x_s| \le \eta_x$, $|\delta u_s| \le \eta_u$, and $\eta_x < |x_s|$. The estimated ratio obeys*

$$|\widehat{p} - p| \le \frac{\eta_u + |p|\eta_x}{|x_s| - \eta_x}. \tag{41}$$

For an interior masked interval $[a, b]$ of label width w, suppose $p \in C^2([a, b])$, $\|p''\|_\infty \le M$, and the endpoint ratio errors are at most E. Linear interpolation then satisfies $\|\widehat{p} - p\|_\infty \le E + Mw^2/8$ on $[a, b]$. For an endpoint gap filled by a constant extension, the corresponding bound is $E + w\|p'\|_\infty$.

Proof. Apply the triangle inequality and $|x_s + \delta x_s| \ge |x_s| - \eta_x$ to Equation (40). The interpolation of endpoint errors is a convex combination bounded by E. The interpolation remainder is bounded by $M(s - a)(b - s)/2 \le Mw^2/8$; the constant-extension bound is the mean value theorem. □

Here derivative errors include sampling, noise, and smoothing bias. On a uniform grid of spacing Δs, bounded coordinate noise ϵ contributes at most $\epsilon/\Delta s$ to centered differences and $2\epsilon/\Delta s$ at the endpoints. Gaussian noise has no deterministic amplitude bound; the proposition is conditional on its realized derivative errors. It supplies no uniform cusp bound or bound on the subsequent learned inverse's condition number.

H.3 Scope of the two-dimensional experiment

For a two-parameter Legendrian surface $s = (s_1, s_2)$, the same identity becomes $D_s U = P\, D_s X$. Where the 2×2 matrix $D_s X$ is invertible, this determines $P = D_s U (D_s X)^{-1}$. At a rank-deficient point, the linear equation alone leaves components of P undetermined. Continuity from neighboring regular points may constrain them, but the one-dimensional simple-zero argument gives neither a general uniqueness statement nor a stable reconstruction procedure for a sampled caustic surface. Accordingly, the two-dimensional inverse benchmark supplies terminal momentum and tests inversion from the full jet. It does not test lifting a projected (x, u) surface.

I Additional Discussion and Scope

Geometry, inversion, and prediction. The three properties tested here answer different questions. Contactness means that a map preserves the hyperplane field $\ker \alpha$ wherever its contact factor is

defined and nonzero. Invertibility means that a state has a unique preimage on the domain of a particular map. Accurate prediction asks whether the fitted map approximates the reference flow on states visited during rollout. The RealNVP comparison in Table 16 separates the first two properties: an invertible model can have a substantially larger Legendrian residual. The oscillator results in Table 7 separate geometry from prediction: a contact map can have a small sampled pullback defect and still accumulate appreciable trajectory error.

The physical inverse adds a conditioning question. If a terminal reference state differs from a state on the learned forward orbit by δz_T, then a local inverse propagates that discrepancy through derivatives of Φ_θ^{-1}. A model self-cycle measures algebraic reversal on model-generated states, whereas a reference-state backcast combines forward-model error with inverse amplification. The perturbation experiment in Appendix G.3 measures local sensitivity; contact structure alone supplies no uniform conditioning bound.

What the diagnostics measure. The oscillator contact defect evaluates a Jacobian pullback identity on a finite cloud. The front diagnostic instead estimates $U_s - PX_s$ in one dimension, or $D_sU - PD_sX$ in two dimensions, by differences along sampled rays. The latter tests whether a predicted front is approximately Legendrian at the grid resolution. It includes discretization error and can be small on a particular front even if a map violates contactness elsewhere. Conversely, a contact map can receive a nonzero sampled front residual. The two diagnostics therefore answer different sampled questions and are reported alongside forward state error; their definitions appear in Appendix E.

The constructions apply to smooth contact maps independently of Hamilton–Jacobi equations. The front experiments are useful because their Legendrian geometry and projected folds make structure and inversion observable in one setting. Applying the same architectures to other contact systems still requires a suitable jet-state representation and a study of the relevant state distribution. In the front inverse task, an ordered full curve supplies information that a selected single-valued solution has discarded. With noisy, unordered, or incomplete observations, the lift and branch-identification problem would require additional assumptions and measurements.

I.1 Additional Related Work

From generators to maps. Hamiltonian neural networks fit a generator and use an integrator to obtain trajectories (Greydanus et al., 2019; Bertalan et al., 2019). Neural ODEs likewise parameterize a vector field (Chen et al., 2018). The numerical integrator is then part of the prediction mechanism. Direct map learning instead fits the sampled evolution operator. It can make inversion an architectural operation, but the learned map must encode the geometric constraint itself if that constraint is desired. This is the setting of the present paper.

For conservative systems, SympNets impose symplectic structure on the learned map (Jin et al., 2020). Hénon-based constructions connect compositions of simple symplectic maps to approximation of richer symplectic dynamics (Turaev, 2003; Burby et al., 2021; Duruisseaux et al., 2023). Our strict atoms retain the same (x, p) Hénon-like update and add the action increment required by the strict contact lift. As shown in Appendix B, C^r approximation of the symplectic map controls the normalized action primitive and the strict lift in C^r for $r \geq 1$; for the C^0 conclusion, we use C^1 symplectic approximation. This extends the conservative map viewpoint without assigning action-dependent feedback to the strict class.

Dissipation and contact maps. GENERIC and metriplectic formulations combine reversible and irreversible terms at the generator level (Grmela & Öttinger, 1997; Zhang et al., 2022). Port-Hamiltonian and dissipative Hamiltonian neural models provide other structured descriptions of open or dissipative dynamics (Bhardwaj & Bajaj, 2026; Sosanya & Greydanus, 2022). Contact Hamiltonian dynamics represents a different geometric route in which dependence on the action coordinate changes the (x, p) evolution (Bravetti, 2017). Contact variational and splitting integrators construct structure-preserving maps from a specified Hamiltonian (Vermeeren et al., 2019; Bravetti et al., 2020); their realization can require implicit solves or tractable subflows. In direct map learning, Testa et al. (2025) use compositions of tractable contact Hamiltonian subflows. We compare the GCF contactomorphism parameterization as a map learner; its broader latent-dynamics, ensemble, and control framework is outside this comparison. Recent conformally symplectic map networks prove fixed-step density for the subclass whose conformal factor is constant across phase space (Gong et al., 2026); general contact flows allow this factor to depend on state. Our factorization uses strict

lifts together with prolongations of invertible base maps. Strict approximation and contact splitting yield a modular universality theorem. Flow maps generated by smooth neural vector fields give one instantiation, and the theorem applies to any base family satisfying the same derivative-level density requirements (Corollary 4.3).

Invertible coupling flows, including RealNVP, offer direct inverses without imposing a contact pullback identity (Dinh et al., 2017; Papamakarios et al., 2021). Their general diffeomorphism approximation results provide complementary expressivity guarantees for invertible maps (Teshima et al., 2020; Ishikawa et al., 2023; Lyu et al., 2022). The RealNVP control in Section 7 also gives an empirical reason to evaluate invertibility and contactness separately.

Fronts and PDE observations. Operator learners such as DeepONet and Fourier neural operators approximate maps between function spaces (Lu et al., 2021; Li et al., 2021). Transport-adapted reduction methods address moving solution features (Rim et al., 2018; Peherstorfer, 2020). Our front experiments instead evolve individual characteristics in jet space and compare their complete lifted states. Classical caustic theory describes why a smooth transported lift can acquire a folded projection (Arnol'd, 1990). A selected viscosity or entropy solution is a different object from that complete projected front (Crandall & Lions, 1983; Evans, 2010). The distinction sets the observation requirements of our inverse experiments: the 1D lift uses all ordered branches, while the 2D example supplies the complete terminal jet.