MLPDEs 2026 · Erlangen, June 22-24

Structure preservation and deep learning for Leraning mechanical systems from Data

Elena Celledoni
Norwegian University of Science and Technology (NTNU)

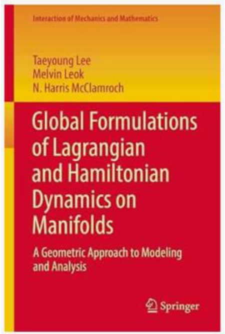
MSCA-SE: REMODEL Project ID: 101131557
aiD funded by RCN

Structure preservation and deep learning

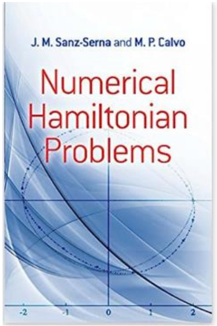
Learning from Position-Only Data

Energy-Behavior Preservation

Stability of Curved Spaces



Geometric numerical integration - Structure preserving algorithms – with data



- In presence of symmetries, invariants and other geometric structure in evolutionary equations, e.g. Hamiltonian systems:

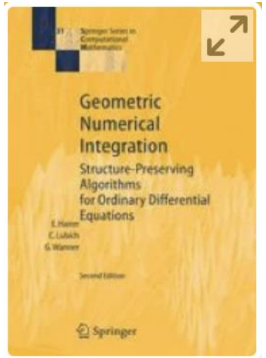
$$\dot{y} = J\nabla H(y).$$

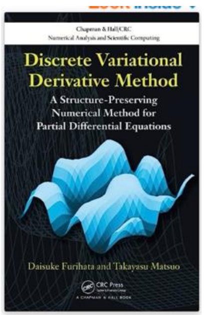
-Design numerical methods that preserve the structure, e.g. symplectic methods.

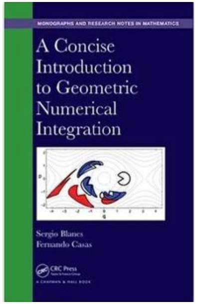
- Backward-Error-Analysis for Hamiltonian systems for ODEs
- For Hamiltonian PDEs this is much more complicated

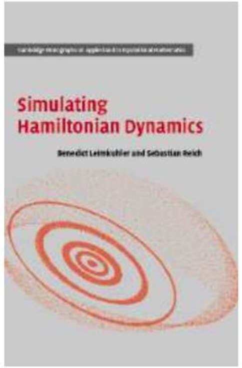
-Lately a pressing demand to feed data into models

Structure-preserving deep learning
E Celledoni, MJ Ehrhardt, C Etmann, RI McLachlan, B Owren, CB Schönlieb, F Sherry
European journal of applied mathematics 32 (5), 888-936









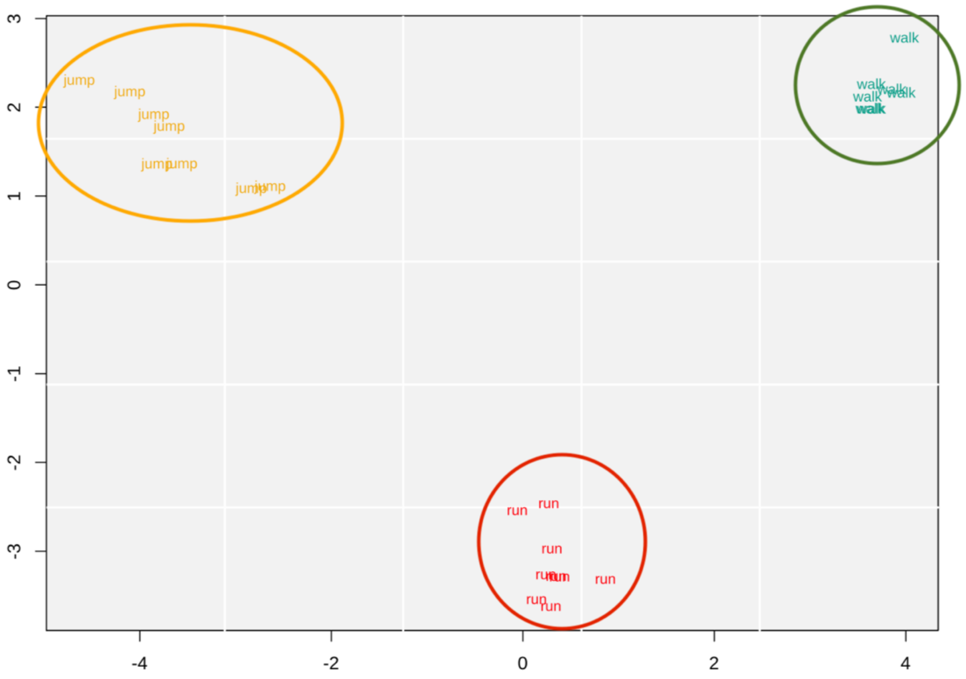
Do structure preservation and geometry matter: Two interesting avenues

- In the dynamical systems view of neural networks: impose structure and geometry on ODE underlying the NN, use a structure preserving discretization to obtain neural network approximations with guaranteed geometric properties.

- Use prior knowledge of geometric structure, invariants and known geometry in the data as inductive bias for solving problems with neural networks (e.g. learning solutions of PDEs or learning dynamical systems from data).

Classifying running, walking, jumping animations as shapes

Celledoni, Eslitzbichler, Schmeding · J. Geometric Mechanics · 2016



Infinite dimensional manifold of curves on $SO(3)^N$ $\quad I = [0, T]$

$$\mathcal{P} := \mathrm{Imm}(\mathrm{I}, SO(3)^N)$$

A *reparametrization invariant* Riemannian distance on $\mathcal{P}$ $\quad d_{\mathcal{P}}$

Inducing a well-defined distance on

$$\mathcal{S} := \mathrm{Imm}(\mathrm{I}, SO(3)^N)/\mathrm{Diff}^+(\mathrm{I})$$

$$d_{\mathcal{S}}([c_0],[c_1]) := \inf_{\varphi \in \mathrm{Diff}^+(\mathrm{I})} d_{\mathcal{P}}(c_0, c_1 \circ \varphi)$$

- $SO(3)^N$ Geometry of rotations is preserved under discretization.
- We treat time curves as shapes: use a *reparametrisation invariant* distance function: $d_{\mathcal{P}}$
- Dynamic programming, to find the *optimal reparametrisation*.

Discrete Forced Lagrangian Neural Networks (DFLNN)

Hansen, Celledoni, Tapley · arXiv:2505.20370

Key idea: use the discrete Lagrange–d'Alembert principle.
Train on position triples (q_n-1, q_n, q_n+1) — velocity is never observed or differentiated.

No velocity/acceleration noise

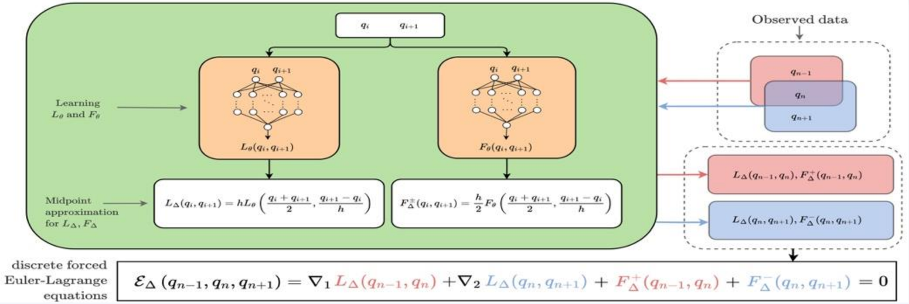
Discrete EL equations on position triples replace the continuous Euler–Lagrange ODEs entirely.

Conservative / forced split

Lagrangian L^d captures conservative dynamics; a separate learned network captures dissipation and control.

Symplectic structure

The variational derivation ensures the learned discrete flow is symplectic — good long-term energy behavior for free.



- S. Ober-Blobaum, C. Offen, Variational Learning of Euler-Lagrange Dynamics from Data. JCAM, 421, 2023.
- C. Offen, S. Ober-Blobaum, Learning discrete Lagrangians for variational PDEs from data and detection of travelling waves, GSI 2023, arXiv:2302.08232.
- Miles Cranmer, et al. Lagrangian neural networks, ICML 2020.
- Xiao, S. et al. Generalized Lagrangian Neural Networks. Comm. In Comp. Phys. (2026).

Lagrange d'Alembert principle - Balance of forces

LdA principle

$$\delta\left(\int_{t_0}^{t_N} L\left(q(t),\dot{q}(t)\right) dt\right) + \int_{t_0}^{t_N} F\left(q(t),\dot{q}(t),u(t)\right)\cdot\delta q(t)\,dt = 0,$$

where δ denotes variations that vanish at the endpoints

$q(t_0) = q_0$ and $q(t_N) = q_N$,

$L : TQ \to \mathbb{R}$ is the Lagrangian function

$F : TQ \to TQ^*$ is the Lagrangian force, a fiber preserving map

$F : (q,\dot{q}) \mapsto (q, F(q,\dot{q}))$,

Euler Lagrange Equations

$$0 = \mathcal{E}(L,F)(q,t) := \frac{\partial L}{\partial q}\left(q(t),\dot{q}(t)\right) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\left(q(t),\dot{q}(t)\right)\right) + F\left(q(t),\dot{q}(t)\right)$$

$$L_\Delta(q_n, q_{n+1}) \approx \int_{t_n}^{t_n+h} L(q(t), \dot{q}(t))\, dt,$$

$$F_\Delta^-(q_n, q_{n+1}, h)\cdot \delta q_n + F_\Delta^+(q_n, q_{n+1}, h)\cdot \delta q_{n+1} \approx \int_{t_n}^{t_n+h} F(q(t), \dot{q}(t))\cdot \delta q(t)\, dt,$$

$L_\Delta : Q\times Q \to \mathbb{R}$ (discrete Lagrangian), $F_\Delta^\pm : Q\times Q \to T^*Q$ (discrete force)

$$\pi_Q^- = \pi_Q \circ F_\Delta^- \qquad \pi_Q^- : Q\times Q \to Q, \qquad \pi_Q^+ = \pi_Q \circ F_\Delta^+ \qquad \pi_Q^+ : Q\times Q \to Q,$$

with $\pi_Q : T^*Q \to Q$ the projection.

Discrete Lagrange d'Alembert principle: using only position data

The discrete Lagrange-d'Alembert principle: find a discrete trajectory $\{q_n\}_{n=1}^N$ s.t.

$$\delta \sum_{n=0}^{N-1} L_\Delta(q_n, q_{n+1}) + \sum_{n=0}^{N-1} \left[F_\Delta^-(q_n, q_{n+1})\cdot \delta q_n + F_\Delta^+(q_n, q_{n+1})\cdot \delta q_{n+1}\right] = 0,$$

for all variations $\{\delta q_n\}_{n=0}^N$ vanishing at the endpoints $\delta q_0 = \delta q_N = 0$.

Discrete forced Euler-Lagrange equations:

$$0 = \mathcal{E}_\Delta(L_\Delta, F_\Delta)(q_{n-1}, q_n, q_{n+1}) := D_2 L_\Delta(q_{n-1}, q_n) + D_1 L_\Delta(q_n, q_{n+1}) + F_\Delta^+(q_{n-1}, q_n) + F_\Delta^-(q_n, q_{n+1}), \quad n = 2, \dots, N-1,$$

where D_1 and D_2 denote differentiation with respect to the first and second variable.

- J E Marsden and M West, *Discrete variational integrators*, Acta Numerica, (2001);
- DM de Diego, R. Sato Martin de Almagro, *Variational order for forced Lagrangian systems*, Nonlinearity, 2018.

Goal: learn NNs $L_\theta \approx L$ and $F_\theta \approx F$ from the observed mechanical trajectories

Inverse discrete LdA setting:
Given multiple observed (discrete) trajectories $\{q_n\}_{n=0}^N$ we seek approximations to the Lagrangian and forces (L, F).

We replace (L, F) with neural networks approximations

$$L_\theta \approx L, \qquad F_\theta \approx F$$

(L_θ, F_θ) are then the unknowns of our problem which we find minimizing the loss function

$$\mathcal{L} = \mathcal{L}_{\text{mech}}(L_\theta, F_\theta) + \mathcal{L}_{\text{reg}}$$

$\mathcal{L}_{\text{mech}}$: residual of discrete Euler Lagrange equations.

$\mathcal{L}_{\text{reg}}$: promotes that L_θ is a regular Lagrangian i.e.

$$S := \left(\frac{\partial^2 L_\theta(q, \dot{q})}{\partial \dot{q}_i \partial \dot{q}_j} \right),$$

is invertible. Impose this constraint in the loss function on N_r regularization points.

DFLNN: Key Results on Real Data

Human motion capture

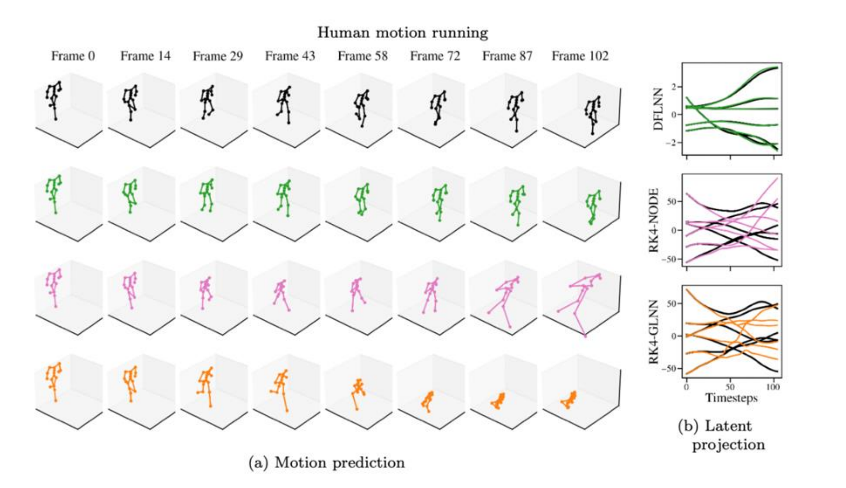
Trained on **position-only** mocap sequences. Reconstructed trajectories match reference; the conservative core generalises beyond training data.

Interpretable decomposition

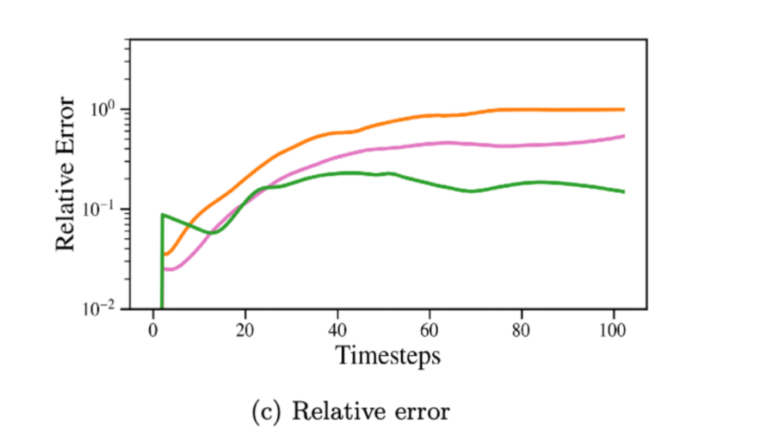
Switch off the learned forced term → **recover the conservative** Lagrangian alone. A black-box model cannot do this.

Independent validation

StrDLNN (IEEE RA-L 2026) builds directly on our discrete forced EL formulation, adding structured mass matrices and **friction identification** on a **UR10e.**

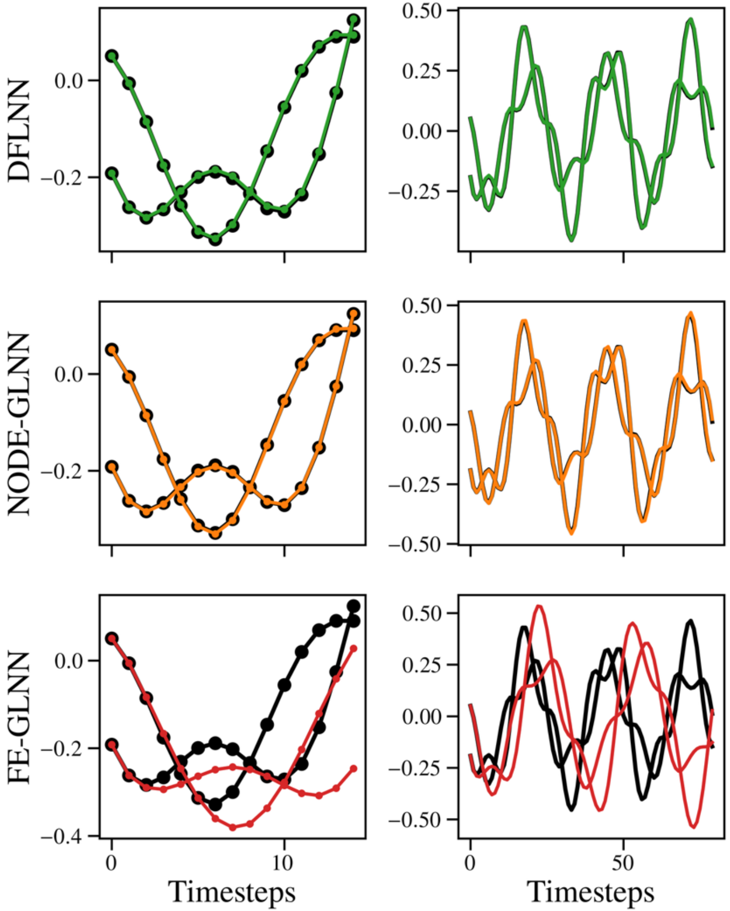


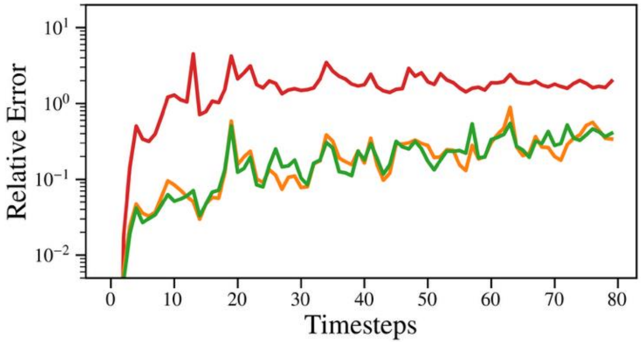
(a) Motion prediction
(b) Latent projection

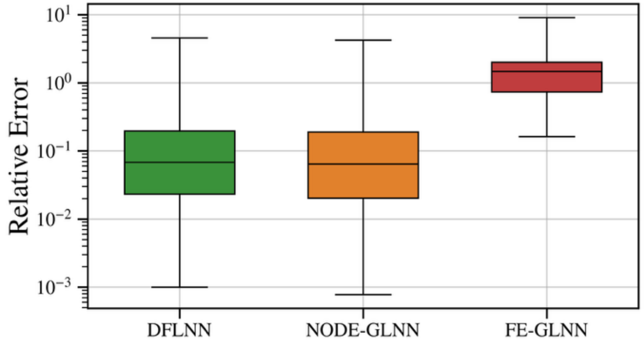


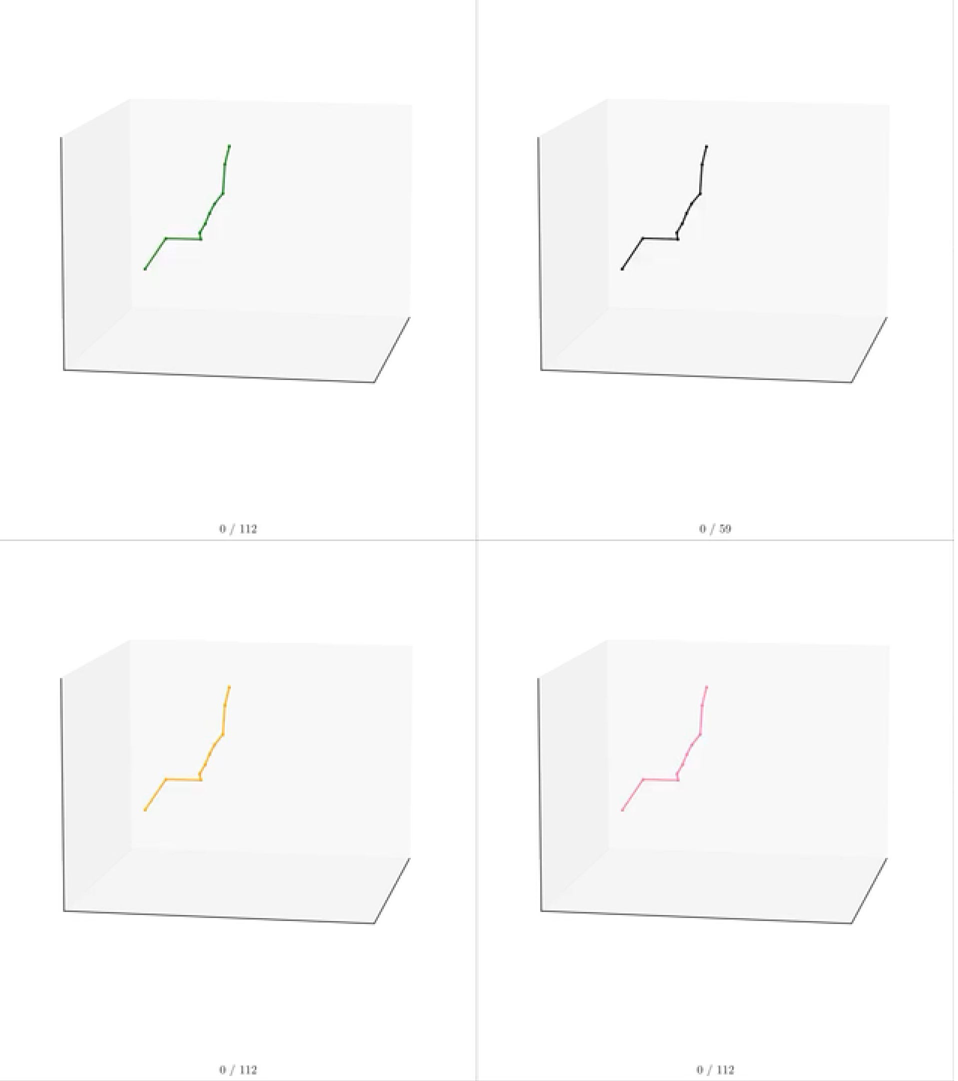
(c) Relative error

Switching off the forces: double pendulum - conservative dynamics









- Green DFLNN
- Black Ground Truth [0,59]
- Yallow GLNN
- Neural ODE

DFLNN lears on time [0,59] repeats the conservative movement on [60,112]

GLNN fails at time 30
Neural ODE fails at time 80.

The Manifold Setting — New Work

Related work

- Duruisseaux, et al., PMLR, 2023.
- Campolo, et al., GSI Conference 2025.

Celledoni, Ghirardelli, Hansen, Martín de Diego, Owren · In preparation

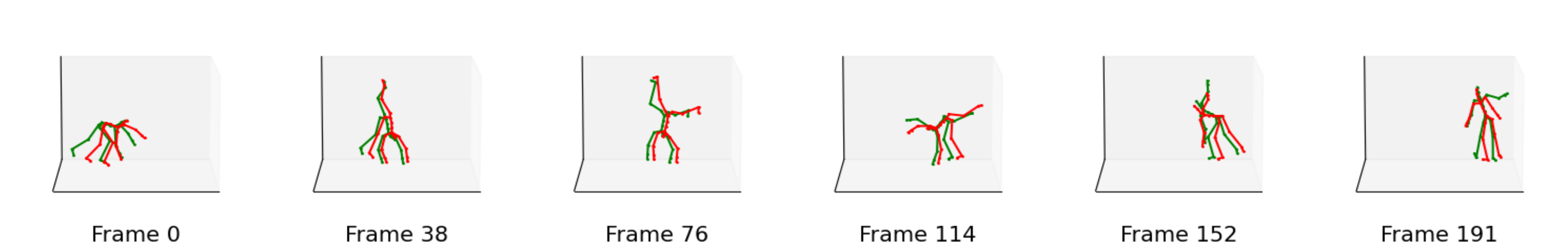
Multibody configuration spaces live on $SO(3)^n \times SE(3) \times$ constraints — never flat Euclidean space.
The discrete forced Lagrangian framework must be extended to this setting.

What we extend

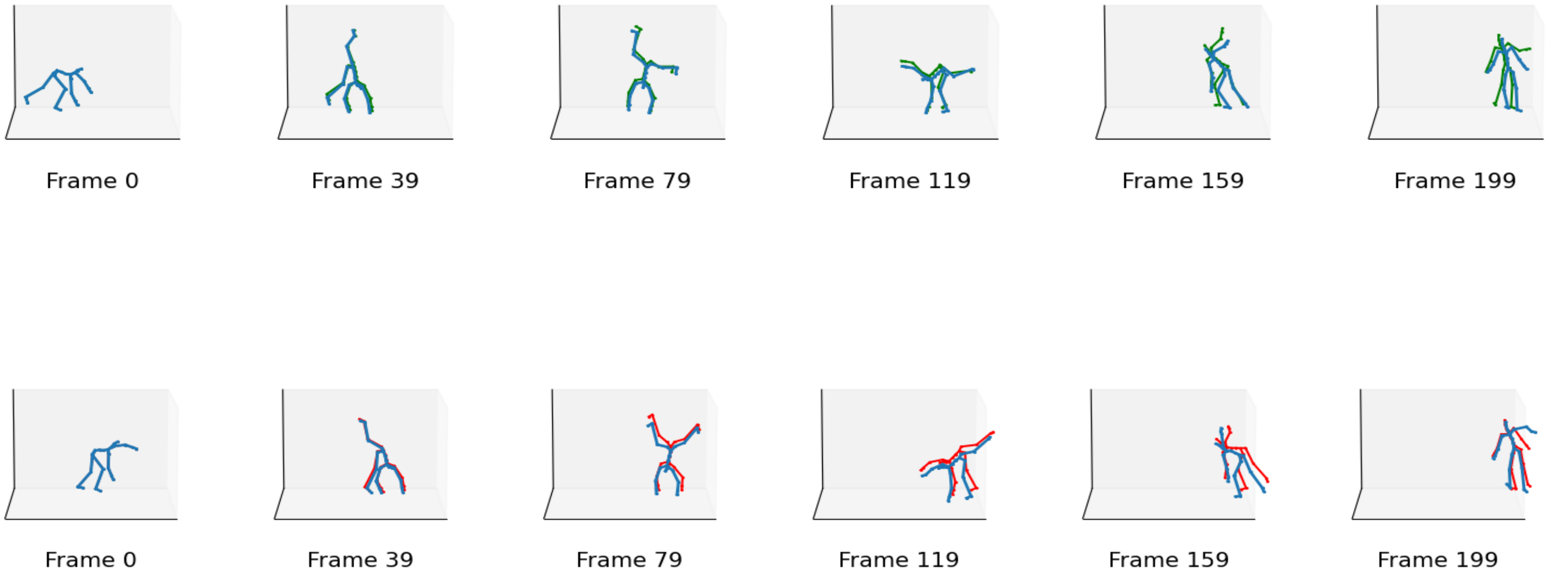
- Discrete Lagrangian $L^d(q_k, q_{k+1})$
- now defined on G × G for a Lie group
- Forced discrete Euler–Lagrange equations on TG
- Retraction-based updates intrinsic to the manifold

Why this matters

- Rigid body links on SO(3)/SE(3)
- Spherical/revolute joints as constraint manifolds
- No Euclidean approximations, no coordinate singularities
- Geometric mechanics rigor



Cartwheel motion: large rotations



Predicted movement given initial poses from the two sequences used during training of the motion across all body segments. Research Square preprint [10.21203/rs.3.rs-10120174/v1](https://doi.org/10.21203/rs.3.rs-10120174/v1).

The PDE Setting — Variational Learning for Hamiltonian PDEs

Celledoni & Jackaman · JCP 2021 · Offen & Ober-Blöbaum · GSI 2023 · arXiv:2302.08232

The discrete Lagrangian framework extends from mechanical systems to Hamiltonian PDEs via multisymplectic structure — the same variational principle, now over space-time.

Multisymplectic FEM (JCP 2021)

- Multisymplectic structure: a closed 2-form over space-time encodes conservation in both space and time simultaneously
- Key result: local energy balance holds cell-by-cell — not just globally, and survives mesh refinement
- Arbitrary-order FEM; applied to wave equation, KdV and nonlinear Schrödinger
- Contrast: standard FD/FV methods typically only enforce global energy balance

Learning Variational PDE Lagrangians

- Offen & Ober-Blöbaum: learn discrete Lagrangian density from PDE data — no explicit equation knowledge needed(GSI 2023)
- Variational structure in space-time: discrete d'Alembert principle on a space-time lattice (lifted from DFLNN to fields)
- Travelling-wave solutions detected; structure encoded in learned L^d
- Natural generalisation: replace mechanical $L^d(q_k,q_{k+1})$ with field density $L^d(u_{ij}, u_{i+1j}, u_{ij+1})$

Open question: Can DFLNN-type variational learning be lifted to space-time — learning multisymplectic Lagrangian densities from data while preserving the local conservation law cell-by-cell? Other PDEs: geometrically exact beams and shells; geometric PDEs; several space variables.

Euler Elastica — A Constrained Geometric Variational Problem

Celledoni, Çokaj, Leone, Leyendecker, Murari, Owren, Sato Martín de Almagro, Stavole · CMAME 2024

Euler's elastica: minimise $\int \kappa^2 \, ds$ subject to the arc-length constraint — a geometric variational PDE on planar curves, connecting shape analysis, constrained mechanics, and learned approximation.

Discrete approach

- Train on discrete equilibrium configurations for varied boundary conditions
- Network predicts discrete solutions for unseen BCs; arc-length and clamping constraints built into architecture
- Learned map: BC → equilibrium shape, without re-solving the variational problem
- Connection to DFLNN: same discrete Lagrangian philosophy, now on an infinite-dimensional configuration space of curves

Continuous approach

- Same discrete training data; network outputs a continuous curve rather than discrete points
- Enforces smoothness and constraint satisfaction directly in the loss
- Both approaches validated across a range of clamping angles and lengths
- Opens the door to structure-preserving learning for constrained geometric PDEs: Willmore flow, curve-shortening flow, elastic surfaces

Thread: shape analysis (Part I) → constrained variational learning (elastica) → DFLNN on manifolds (Part II) → multisymplectic fields (slide 13). Geometry and structure at every level.

When incorporating data: What should we preserve?

Symplecticity OR Time-symmetry and energy (resp. energy dissipation)

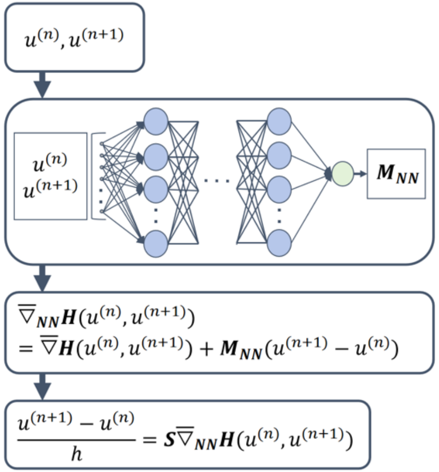
Structure-preserving deep learning
E Celledoni, MJ Ehrhardt, C Etmann, RI McLachlan, B Owren, CB Schönlieb, F Sherry
European journal of applied mathematics 32 (5), 888-936

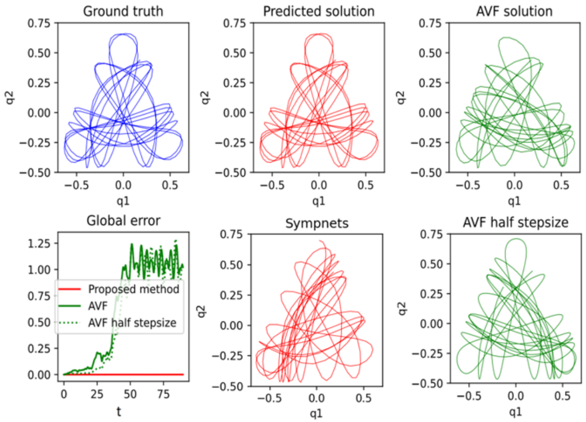
Dynamical systems–based neural networks
E Celledoni, D Murari, B Owren, CB Schönlieb, F Sherry
SIAM Journal on Scientific Computing 45 (6), A3071-A3094

Equivariant neural networks for inverse problems
E Celledoni, MJ Ehrhardt, C Etmann, B Owren, CB Schönlieb, F Sherry
Inverse Problems, 2021

Energy-preserving integrators applied to nonholonomic systems
E Celledoni, M Farré Puiggalí, EH Høiseth, D Martin de Diego
Journal of Nonlinear Science 29 (4), 1523-1562, 2019

Universal Energy Preserving resp. Dissipative Discrete Gradient Integrators





- Discrete gradients are a simple approach for numerical discretization of $\dot{u} = S\,\nabla H(u)$ with S skew-symmetric (energy preserving) or S negative definite (energy dissipative).
- A discrete gradient $\bar{\nabla} H(u^n, u^{n+1})$ satisfies $\bar{\nabla} H(u, u) = \nabla H(u)$ and

$$(u^{n+1} - u^n)^T \bar{\nabla} H(u^n, u^{n+1}) = H(u^{n+1}) - H(u^n)$$

- Characterization of all discrete gradients:

$$\bar{\nabla} H(u^n, u^{n+1}) = \bar{\nabla}_0 H(u^n, u^{n+1}) + M(u^n, u^{n+1})(u^{n+1} - u^n)$$

Celledoni, Owren, Shen, Xu, Yaguchi · NeurIPS 2025

Residual networks can be built by discretizing ODEs

$$\dot{y} = \sigma(A(t)y(t) + b(t)), \qquad y(0) = x, \qquad t \in [0, h]$$

we get

$$\varphi_{\theta_\ell} = \mathrm{id} + hX_{\theta_\ell}, \quad X_{\theta_\ell} : x \mapsto \sigma(A_\ell x + b_\ell), \quad \theta_\ell := (A_\ell, b_\ell)$$

We optimize an appropriate loss function to determine the parameters

$$\min_{\varphi = \varphi_{\theta_L} \circ \cdots \circ \varphi_{\theta_1}} E(\varphi) = \min_{\{\theta_\ell\}_{\ell=1}^{L}} E(\varphi_{\theta_L} \circ \cdots \circ \varphi_{\theta_1})$$

We can build new architectures by deciding the features of the underlying ODE and then discretizing with a (structure preserving) numerical integrator.

- Symplectic networks: built using symplectic discretizations of Hamiltonian systems
- We can study stability and impose L-Lipschiztness of NNs: conditional contractivity of the numerical integrators is useful for this.

CNNs as PDE Discretizations

Celledoni, Jackaman, Murari, Owren · Journal of Computational Physics 538 (2025) 114166

Two-layer CNNs with appropriate activations can exactly represent any finite difference discretization of a linear PDE via the method of lines — connecting network architecture directly to numerical PDE structure.

Architecture = FD Stencil

- Convolutional weights ⟷ finite difference coefficients
- Bias terms ⟷ boundary contributions
- Activation function ⟷ temporal integrator step
- Validated on advection, heat, and Fisher–KPP equations

Stability Through Structure

- Norm-preserving activations enforce stability of the encoded discretisation
- Small networks suffice: FD consistency is exact, not approximate
- Overfitting avoided by construction — the inductive bias is the PDE

Bridge to UEPI

- Once architecture encodes a PDE discretisation, ask: is it energy-preserving?
- UEPI (NeurIPS 2025) learns the discrete gradient that makes it so
- Architecture design by structure, not by chance

→ *See also: preceding slide (ResNets from ODEs) for the temporal/ODE analogue (Haber & Ruthotto 2018 line of work)*

Nonlinear Stability on Riemannian Manifolds

- *Arnold, Celledoni, Cokay, Owren, Tumiotto — B-stability on manifolds. J. Comp. Dyn. 2025.*
- Ghirardelli, Owren, Celledoni — Conditional stability. J FoCM 2026 to appear.

The connection: ResNets (preceding slide) encode explicit Euler on an ODE — the conditional stability theorem gives step-size bounds for exactly this scheme on Riemannian manifolds. DFLNN on SO(3) uses geodesic implicit Euler — the B-stability theorem applies directly. Parts II–III build learned integrators on manifolds; Part IV provides the theoretical warranty.

B-stability on manifolds JCD 2025

- Generalises Butcher's B-stability to Riemannian manifolds via non-expansive systems
- Geodesic implicit Euler is B-stable on manifolds of non-positive sectional curvature
- Expansiveness can occur on S^2 — positive curvature is dangerous
- Direct relevance: SO(3) has positive curvature. Robotics and multibody systems live here.

$$\langle X(t,y_2)-X(t,y_1), y_2-y_1\rangle \leq -\nu\|y_2-y_1\|^2$$

Conditional stability of explicit Euler JFoCM 2026.

- Generalizes conditional stability results by Dahlquist and Jeltch (circle contractivity) for explicit methods to Riemannian manifolds.
- Derives step-size bounds via manifold-adapted cocoercivity
- Non-zero curvature shrinks stability region — tight bounds on S^n and hyperbolic space
- Direct relevance: SO(3) has positive curvature. Robotics and multibody systems live here.

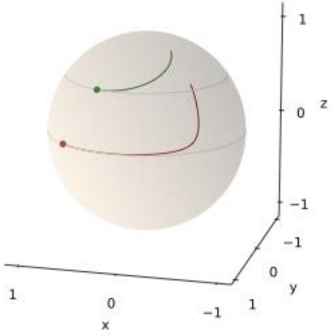
Cocoercivity condition:

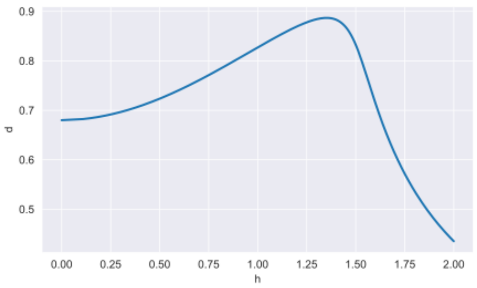
$$\langle X(t,y_2)-X(t,y_1), y_2-y_1\rangle \leq -\bar{\nu}\|X(t,y_2)-X(t,y_1)\|^2, \quad \bar{\nu}\geq 0,$$

Theorem If $\mathcal{M}$ is a Riemannian manifold with non-positive sectional curvature then the geodesic implicit Euler method is B-stable.

Example Space of $n \times n$ symmetric positive definite matrices.

Geodesic Implicit Euler is not B-stable on the sphere. Counterexample.





- Non-expansive vector field (on the northern hemisphere)

$$\dot{y} = X(y) = a \times y, \qquad a = [0, 0, 1].$$

- (Left) One step of GIE applied with increasing step size h, starting from two different initial values.
- (Right) Geodesic distance: $d(y_1, z_1)$ plotted as a function of h, where $y_0 = \exp_{y_1}(-hX(y_1))$, $\quad z_0 = \exp_{z_1}(-hX(z_1))$.

Cocoercivity condition:

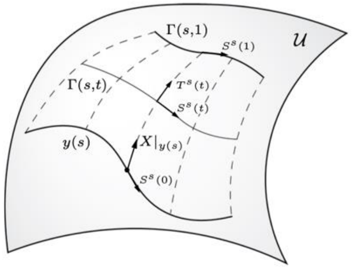
$$\langle X(t,y_2) - X(t,y_1), y_2 - y_1 \rangle \le -\bar{\nu} \| X(t,y_2) - X(t,y_1) \|^2, \quad \bar{\nu} \ge 0,$$

- Consider the variation

$$\Gamma(s,t) = \exp_{y(s)}(t\, h\, X|_{y(s)})$$

- We look at the vector fields $T^s = \partial_t \Gamma$, $S^s = \partial_s \Gamma$, S^s is a *Jacobi field* along the geodesic $t \mapsto \Gamma(s,t)$ and it is enough to show that

$$\| S^s(1) \|^2 - \| S^s(0) \|^2 \le 0$$



α-cocoercivity $\alpha > 0$,

$$\langle \nabla_{v_p} X, v_p \rangle \le -\alpha \| \nabla_{v_p} X \|^2$$

Open Problems

Well-posedness of learned PDEs

Does the learned equation admit a well-posed IVP? Is it hyperbolic, parabolic, or elliptic? Learned equations carry no such guarantee — a fundamental gap between data fit and mathematical validity.

Non-smooth phenomena

Mechanical impacts and PDE shocks share the same pathology: the variational/Lagrangian structure breaks at discontinuities. Hybrid variational integrators for impacts; structure-preserving schemes for conservation laws with shocks — both are open frontiers.

Noise in Positions

The discrete EL residual assumes exact q_k. A rigorous probabilistic treatment is missing.

Identifiability

Discrete Lagrangians are defined up to equivalence classes. What exactly does the network learn?

Convergence Guarantees

Does the learned L^d converge to the true L as data grows? Backward-error analysis gives partial answers.

Learning PDE operators

Lifting DFLNN and multisymplectic learning from finite-dimensional manifolds to function spaces: operator learning with variational structure, à la DeepONet/FNO but with built-in local conservation laws.

Thank you

Variational structure — from discrete mechanics to multisymplectic PDEs — is the inductive bias that makes learned dynamical systems trustworthy, and numerical analysis on curved spaces tells us when to trust them.

References

- Celledoni, Eslitzbichler, Schmeding — Shape analysis on Lie groups. J. Geometric Mechanics, 2016.
- Hansen, Celledoni, Tapley — DFLNN. arXiv:2505.20370.
- Hansen, Ghirardelli, Celledoni, Martín de Diego, Owren — Manifold setting. Research Square preprint [10.21203/rs.3.rs-10120174/v1](10.21203/rs.3.rs-10120174/v1).
- Celledoni, Owren, Shen, Xu, Yaguchi — UEPI. NeurIPS 2025.
- Arnold, Celledoni, Cokay, Owren, Tumiotto — B-stability on manifolds. J. Comp. Dyn. 2025.
- Ghirardelli, Owren, Celledoni — Conditional stability. J FoCM 2026 to appear.