Geometric Methods for Adjoint Systems

Brian Tran

Joint work with Prof. Melvin Leok

CGM Research Seminar

This material is based upon work supported by the NSF Graduate Research Fellowship DGE-2038238, by NSF under grants DMS-1813635, DMS-1345013, DMS-2217293, and by AFOSR under grant FA9550-18-1-0288.

Overview

- Motivation
- Background Material
  - Sensitivity Analysis for ODEs
  - Adjoint Sensitivity Analysis for ODEs
- Geometric Integration of Adjoint DAE Systems
  - The Geometry of Adjoint ODE Systems
    - Variational Characterization of Adjoint ODE Systems
  - Adjoint Systems for DAEs
    - Differential-algebraic Equations
    - The Geometry of Adjoint DAE Systems
    - Index Reduction and the Presymplectic Constraint Algorithm
  - Discretization of Adjoint DAE Systems
  - Numerical Example
- Future Directions: Adjoint Systems for Evolution PDEs

Motivation

- Goal: compute the sensitivity of objective function
  - Terminal cost $C(q(t_f))$
  - Running cost $\int_0^{t_f} L(q(t))dt$

  subject to ODE (or DAE) $F(t, q(t), \dot{q}(t)) = 0, q^i(0) = q_0^i$

  w.r.t. perturbation $\delta q^i(0)$
  - Purpose: given an initial state $q^i(0) = q_0^i$, want to know ascent directions
- Examples:
  - Optimization and design
  - Optimal control
  - Neural networks
- Investigate geometric structure to utilize geometric integration to numerically integrate such problems in a structure-preserving manner

Sensitivity Analysis for ODEs

- Terminal cost $C(q(t_f))$
- ODE $\dot{q} = f(q), q(0) = q_0$
  - How does the cost change as we perturb by δq_0
- The direct method:
  - The variational equation

$$\frac{d}{dt}\delta q(t) = Df(q(t))\delta q(t), \quad \delta q(0) = \delta q_0$$

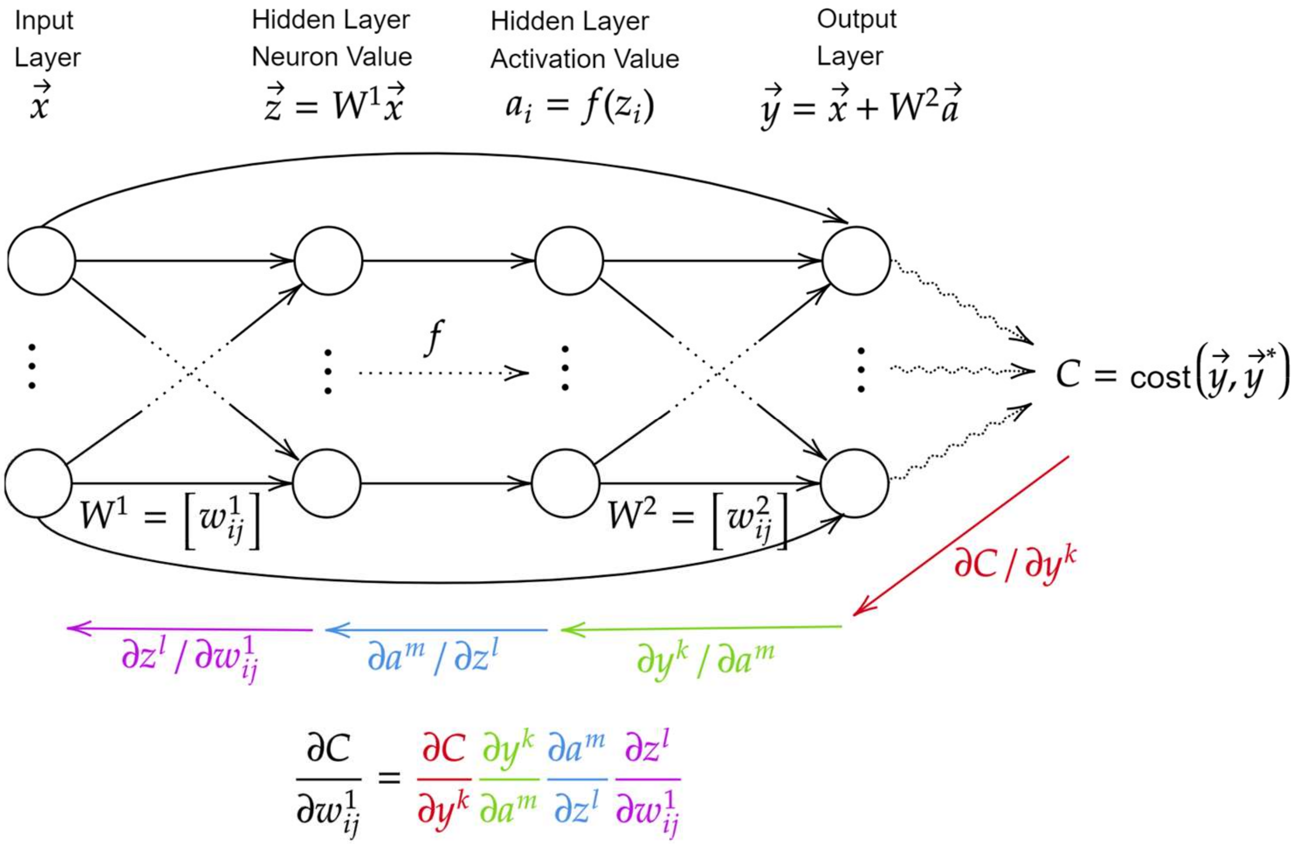
  - The sensitivity is implicitly given by the change $\langle \nabla_q C(q(t_f)), \delta q(t_f) \rangle$ induced by δq_0
  - Drawback: requires $\mathcal{O}(N)$ integrations of the variational equations (N is the dimension of the system)
    - ➢ Prohibitively expensive when N is large

Adjoint Sensitivity Analysis for ODEs

- The adjoint equation associated to the ODE
$$\dot{p} = -[Df(q)]^* p$$
- Given curves satisfying the variational and adjoint equations,
$$\frac{d}{dt}\langle p, \delta q\rangle = -\langle [Df(q)]^* p, \delta q\rangle + \langle p, Df(q)\delta q\rangle = 0$$
  - Gives adjoint conservation law $\langle p(0), \delta q(0)\rangle = \langle p(t_f), \delta q(t_f)\rangle$
- Set $p(t_f) = \nabla_q C(q(t_f))$, the sensitivity is $\langle p(0), \delta q(0)\rangle$
  - Requires only $\mathcal{O}(1)$ backwards integrations or "backpropagations"
  - Drawback: requires $\mathcal{O}(N_C)$ integrations
    (N_C is number of cost functions)
  - Adjoint method advantageous when $N >> N_C$
  - For example, in PDE-constrained optimization, the ODE arises as a semi-discretization of a PDE and the number of cost functions is fixed, while N increases as the semi-discretization is refined

Adjoint Sensitivity Analysis for ODEs …

- Another example: backpropagation for training a neural network



- Neural ODEs: Can view the sequence of layers of a neural network as a discretization of an ODE [6]

$$x_{t+1} = x_t + g(t, x_t, W^{(t)})$$

$$\frac{dx}{dt} = g(t, x(t), W(t))$$

- The sensitivity of the cost with respect to the weights of the neural network obey a discretization of the adjoint equation

The Geometry of Adjoint ODE Systems

- The adjoint system $\begin{aligned}\dot{q} &= f(q), \\ \dot{p} &= -[Df(q)]^* p\end{aligned}$ is a Hamiltonian system
- ODE $\dot{q} = f(q)$ specified by a vector field f over a manifold M
- Define a Hamiltonian and use canonical symplectic form

$$H : T^*M \to \mathbb{R}, \quad H(q,p) = \langle p, f(q) \rangle, \quad \Omega = dq \wedge dp \in \Lambda^2(T^*M)$$

  - The adjoint system is given by Hamilton's equations

$$i_{\widehat{f}}\Omega = dH$$

  - In coordinates,

$$\dot{q} = \partial H/\partial p = f(q), \dot{p} = -\partial H/\partial q = -[Df(q)]^* p$$

  - The adjoint system covers the original ODE
  - Equivalently, the vector field $\widehat{f}$ is a lift of the vector field f, the "cotangent lift"
- In [1], we show that the adjoint conservation law arises from symplecticity of the Hamiltonian flow of the above system. This shows that symplectic integrators are suitable for adjoint sensitivity analysis.

Variational Characterization of Adjoint ODE Systems

- In [1], we develop an intrinsic Type II variational principle for adjoint systems
- Action
$$S[q,p] = \int_0^{t_f} \left(\langle p, \dot{q}\rangle - H(q,p)\right) dt = \int_0^{t_f} \langle p, \dot{q} - f(q)\rangle dt$$
- Goal: derive the adjoint system from a variational principle $\delta S = 0$
- Type I boundary conditions:
$$q(0) = q_0, q(t_f) = q_1 \implies \delta q(0) = 0, \delta q(t_f) = 0$$
  - Do not make sense for adjoint systems, because they cover an ODE
- Type II boundary conditions:
$$q(0) = q_0, p(t_f) = p_1 \implies \delta q(0) = 0, \delta p(t_f) = 0$$
  - For general Hamiltonian systems, Type II boundary conditions do not make intrinsic sense (cannot specify a covector without specifying the basepoint)
  - However, it does make sense if the Hamiltonian systems covers an ODE, since we know to specify p_1 at $q_1 = \Phi_{t_f}(q_0)$
  - Locally, every Hamiltonian system on a cotangent bundle which covers an ODE on the base manifold is an adjoint system

Variational Characterization of Adjoint ODE Systems …

- Type II variational characterization tells us:
  - The boundary conditions used for adjoint sensitivity analysis make intrinsic sense

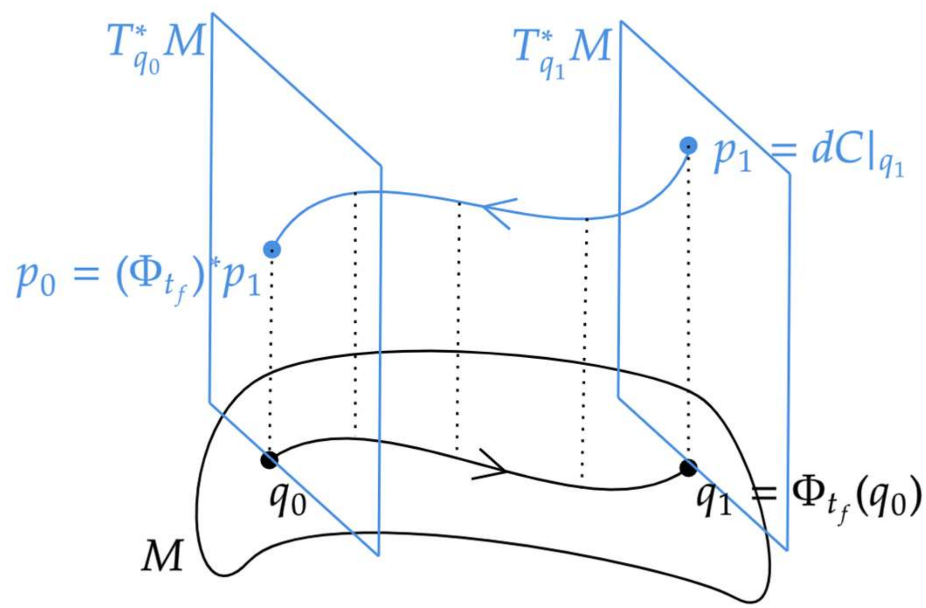
$$q(0) = q_0, p(t_f) = dC(q(t_f))\Big|_{q(t_f)=\Phi_{t_f}(q_0)}$$

  - Gives an intrinsic theoretical justification for the backpropagation used in adjoint sensitivity analysis. Particularly for optimization on
    - Nonlinear spaces
    - Infinite-dimensional spaces, e.g.,

$$\dot{q} = \Delta q$$
$$\dot{p} = -\Delta p$$

  - Can construct structure-preserving integrators for these systems using variational integrators

$$\Phi_t = \text{time } t \text{ flow of } f$$
$$C : M \to \mathbb{R}$$



Differential-Algebraic Equations

- Goal: compute sensitivities of cost functions for systems governed by DAEs
- Generalize adjoint systems to DAEs $\dot{q} = f(q, u),$
$$0 = \phi(q, u)$$
- Dynamic and algebraic variables $q \in M_d, u \in M_a$
- Let $\overline{TM}_d \to M_d \times M_a$ be the pullback bundle of $TM_d \to M_d$ by $M_d \times M_a \to M_d$ and similarly for the cotangent bundle
- Let $\Phi \to M_d \times M_a$ be a vector bundle and Φ^* its dual
- A DAE is specified by sections $f \in \Gamma(\overline{TM}_d), \phi \in \Gamma(\Phi)$
- Define the adjoint system as a presymplectic Hamiltonian system on $\overline{T^*M}_d \oplus \Phi^*$

$$\begin{tikzcd}
(q,u,p,\lambda) \in \overline{T^*M}_d \oplus \Phi^* \arrow[r] \arrow[d] & (q,u,\lambda) \in \Phi^* \arrow[d] & \Phi \arrow[dl] \\
(q,u,p) \in \overline{T^*M}_d \arrow[r] \arrow[d] & M_d \times M_a \ni (q,u) \arrow[d, "\pi_d"] & \overline{TM}_d \ni (q,u,v) \arrow[l] \arrow[d] \\
(q,p) \in T^*M_d \arrow[r] & M_d \ni q & TM_d \ni (q,v) \arrow[l]
\end{tikzcd}$$

The Geometry of Adjoint DAE Systems

- Let Ω_d be the canonical symplectic form on T^*M_d . Pull back this form by the sequence of maps $\overline{T^*M}_d \oplus \Phi^* \to \overline{T^*M}_d \to T^*M_d$ to obtain a presymplectic form
$$\Omega_0 = dq \wedge dp \in \Lambda^2(\overline{T^*M}_d \oplus \Phi^*)$$
- Define the Hamiltonian
$$H : \overline{T^*M}_d \oplus \Phi^* \to \mathbb{R},$$
$$H(q, u, p, \lambda) = \langle p, f(q, u) \rangle + \langle \lambda, \phi(q, u) \rangle$$
- The adjoint DAE system is the presymplectic Hamiltonian system $i_X \Omega_0 = dH$
- In coordinates,

$$\dot{q} = \frac{\partial H}{\partial p} = f(q, u), \qquad \leftarrow \text{dynamical equation}$$

$$\dot{p} = -\frac{\partial H}{\partial q} = -[D_q f(q, u)]^* p - [D_q \phi(q, u)]^* \lambda, \qquad \leftarrow \text{adjoint of dynamical equation}$$

$$0 = \frac{\partial H}{\partial \lambda} = \phi(q, u), \qquad \leftarrow \text{constraint equation}$$

$$0 = -\frac{\partial H}{\partial u} = -[D_u f(q, u)]^* p - [D_u \phi(q, u)]^* \lambda. \qquad \leftarrow \text{adjoint of constraint equation}$$

Index Reduction and the Presymplectic Constraint Algorithm

- Unlike the symplectic case, solutions are not everywhere defined
  - Defined on a submanifold of $P = \overline{T^*M}_d \oplus \Phi^*$
- Basic idea: constraints require that the vector field X lies on a constraint submanifold of P. In order for solutions, i.e., integral curves of X, to stay on the submanifold, X must be tangent to the submanifold where it is defined. The process of obtaining this final constraint submanifold to which X is tangent is known as the presymplectic constraint algorithm
- Related idea: for the base DAE, the index is the number of differentiations of the constraints needed to obtain an ODE
  - For example, index 1: $\dot{q} = f(q,u),$

    $$0 = D_q\phi(q,u)\dot{q} + D_u\phi(q,u)\dot{u}$$
  - ODE in $(\dot{q},\dot{u})$ if $D_u\phi(q,u)$ is invertible wherever $\phi(q,u) = 0$
  - Alternatively, by the implicit function theorem, $\dot{q} = f(q,u(q))$
- Question: can we relate the presymplectic constraint algorithm of the adjoint DAE system to the index of the base DAE?

Index Reduction and the Presymplectic Constraint Algorithm …

- Question: can we relate the presymplectic constraint algorithm of the adjoint DAE system to the index of the base DAE?
- In [1], we show that the presymplectic constraint algorithm for the adjoint DAE system terminates after the number of steps given by the index of the base DAE
- Furthermore, we show that index reduction and forming the adjoint system commute
  - Reduce DAE; form adjoint ODE system
  - Form adjoint DAE system; reduce through presymplectic constraint algorithm
- Can derive results for the adjoint DAE system using the reduced adjoint ODE system; e.g., the adjoint conservation law

$$\langle p(0), \delta q(0) \rangle = \langle p(t_f), \delta q(t_f) \rangle$$

  - This conservation law can also be interpreted as presymplecticity

$$\Omega_0 = dq \wedge dp$$

Discretization of Adjoint DAE Systems

- We extend the Galerkin Hamiltonian variational integrator construction of Leok and Zhang [2] to the setting of presymplectic Hamiltonian systems
- The basic ingredients
  - Finite-dimensional function space approximating curves on P
  - Quadrature rule
  - Enforce discrete Type II variational principle
    - Using a function space which interpolates the quadrature nodes results in a constrained partitioned Runge–Kutta method

$$
\begin{aligned}
q_1 &= q_0 + \Delta t \sum_i b_i f(Q^i, U^i), \\
Q^i &= q_0 + \Delta t \sum_j a_{ij} f(Q^j, U^j), \\
p_1 &= p_0 - \Delta t \sum_i b_i \left([D_q f(Q^i, U^i)]^* P^i + [D_q \phi(Q^i, U^i)]^* \Lambda^i \right), \\
P^i &= p_0 - \Delta t \sum_j \tilde{a}_{ij} \left([D_q f(Q^j, U^j)]^* P^j + [D_q \phi(Q^j, U^j)]^* \Lambda^j \right), \\
0 &= \phi(Q^i, U^i), \\
0 &= [D_u f(Q^i, U^i)]^* P^i + [D_u \phi(Q^i, U^i)]^* \Lambda^i,
\end{aligned}
$$

Discretization of Adjoint DAE Systems …

- Integrator is presymplectic, $\langle p_0, \delta q_0 \rangle = \langle p_1, \delta q_1 \rangle$
- Integrator is "natural": discretization, reduction, and forming the adjoint commute

$$
\begin{tikzcd}
\text{Index 1 DAE} \arrow[rr, "\text{Reduce}"] \arrow[dr, "\text{Discretize}"'] \arrow[dd, "\text{Adjoint}"'] & & \text{ODE} \arrow[dr, "\text{Discretize}"] \arrow[dd, "\text{Adjoint}"] & \\
& \text{Discrete DAE} \arrow[rr, "\text{Reduce}"] \arrow[dd, "\text{Adjoint}"] & & \text{Discrete ODE} \arrow[dd, "\text{Adjoint}"] \\
\substack{\text{Presymplectic Adjoint}\\ \text{DAE System}} \arrow[rr, "\text{Reduce}"] \arrow[dr, "\text{Discretize}"'] & & \substack{\text{Symplectic Adjoint}\\ \text{ODE System}} \arrow[dr, "\text{Discretize}"'] & \\
& \substack{\text{Presymplectic Galerkin}\\ \text{Hamiltonian Variational}\\ \text{Integrator}} \arrow[rr, "\text{Reduce}"'] & & \substack{\text{Symplectic Galerkin}\\ \text{Hamiltonian Variational}\\ \text{Integrator}}
\end{tikzcd}
$$

Discretization of Adjoint DAE Systems …

- As a consequence of this naturality, we prove a variational error analysis result:

Proposition 3.4. *Suppose the discrete generating function $H_d^+(q_0, p_1; \Delta t)$ for the presymplectic variational integrator approximates the exact discrete generating function $H_d^{+,E}(q_0, p_1; \Delta t)$ to order r, i.e.,*

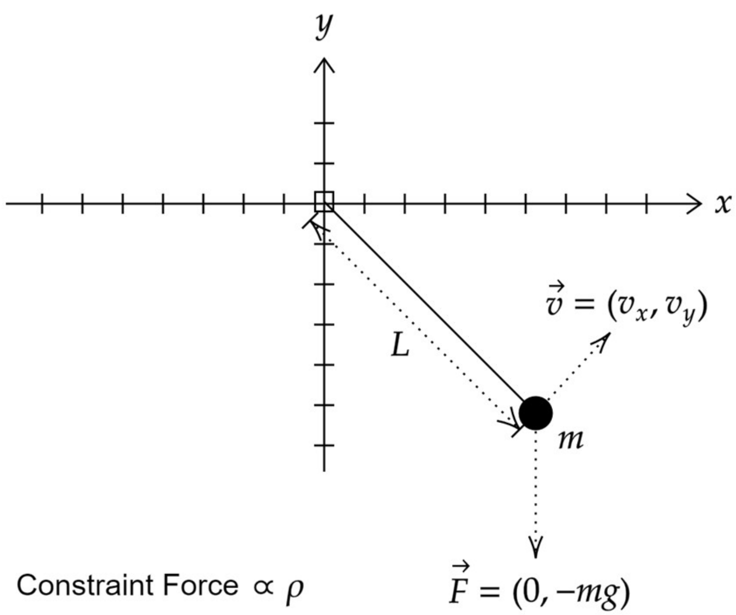
$$H_d^+(q_0, p_1; \Delta t) = H_d^{+,E}(q_0, p_1; \Delta t) + \mathcal{O}(\Delta t^{r+1}),$$

and the Hamiltonian H is continuously differentiable, then the Type II map $(q_0, p_1) \mapsto (q_1, p_0)$ and the evolution map $(q_0, p_0) \mapsto (q_1, p_1)$ are order-r accurate.

- Proof sketch:
  - Use naturality to relate the discretized adjoint DAE system to the discretized and reduced adjoint ODE system
  - Subsequently, apply the variational error analysis result in the ODE case (Schmitt and Leok [3])

Numerical Example

- As the adjoint conservation law arises from presymplecticity, perform a simple example to numerically illustrate that the aforementioned integrator is presymplectic (and hence, suitable for adjoint sensitivity analysis)
- As an academic example, we consider the planar pendulum, as an index 1 DAE:



$$
\begin{aligned}
\dot{x} &= v_x, \\
\dot{v}_x &= \rho x/m, \\
0 &= x^2 + y^2 - L^2, \\
0 &= v_x x + v_y y, \\
0 &= m(v_x^2 + v_y^2) - mgy + L^2 \rho.
\end{aligned}
$$

Numerical Example …

- In terms of the notation we used for adjoint DAE systems: $q = (x, v_x),\ u = (y, v_y, \rho)$

$$f(q,u) = \begin{pmatrix} v_x \\ \rho x/m \end{pmatrix},$$

$$\phi(q,u) = \begin{pmatrix} x^2 + y^2 - L^2 \\ v_x x + v_y y \\ m(v_x^2 + v_y^2) - mgy + L^2\rho \end{pmatrix}$$

$$H(q,u,p,\lambda) = \langle p, f(q,u)\rangle + \langle \lambda, \phi(q,u)\rangle$$

$$= (p_x\ p_{v_x}) \begin{pmatrix} v_x \\ \rho x \end{pmatrix} + (\lambda_1\ \lambda_2\ \lambda_3) \begin{pmatrix} x^2 + y^2 - L^2 \\ v_x x + v_y y \\ m(v_x^2 + v_y^2) - mgy + L^2\rho \end{pmatrix}$$

$$\Omega_0 = dx \wedge dp_x + dv_x \wedge dp_{v_x}$$

$$D_q f(q,u) = \begin{pmatrix} 0 & 1 \\ \rho/m & 0 \end{pmatrix}, \qquad D_q\phi(q,u) = \begin{pmatrix} 2x & 0 \\ v_x & x \\ 0 & 2mv_x \end{pmatrix},$$

$$D_u f(q,u) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & x/m \end{pmatrix}, \qquad D_u\phi(q,u) = \begin{pmatrix} 2y & 0 & 0 \\ v_y & y & 0 \\ -mg & 2mv_y & L^2 \end{pmatrix}$$

Numerical Example …

- The corresponding adjoint DAE system is

$$
\begin{aligned}
\dot{q} &= \frac{\partial H}{\partial p} = f(q,u),\\
\dot{p} &= -\frac{\partial H}{\partial q} = -[D_q f(q,u)]^* p - [D_q \phi(q,u)]^* \lambda,\\
0 &= \frac{\partial H}{\partial \lambda} = \phi(q,u),\\
0 &= -\frac{\partial H}{\partial u} = -[D_u f(q,u)]^* p - [D_u \phi(q,u)]^* \lambda.
\end{aligned}
\quad \Longleftrightarrow \quad
\begin{aligned}
\frac{d}{dt}\begin{pmatrix} x \\ v_x \end{pmatrix} &= \begin{pmatrix} v_x \\ \rho x/m \end{pmatrix},\\
\frac{d}{dt}\begin{pmatrix} p_x \\ p_{v_x} \end{pmatrix} &= -\begin{pmatrix} 0 & 1 \\ \rho/m & 0 \end{pmatrix}^T \begin{pmatrix} p_x \\ p_{v_x} \end{pmatrix} - \begin{pmatrix} 2x & 0 \\ v_x & x \\ 0 & 2mv_x \end{pmatrix}^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix},\\
0 &= \begin{pmatrix} x^2 + y^2 - L^2 \\ v_x x + v_y y \\ m(v_x^2 + v_y^2) - mgy + L^2 \rho \end{pmatrix},\\
0 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & x/m \end{pmatrix}^T \begin{pmatrix} p_x \\ p_{v_x} \end{pmatrix} + \begin{pmatrix} 2y & 0 & 0 \\ v_y & y & 0 \\ -mg & 2mv_y & L^2 \end{pmatrix}^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}
\end{aligned}
$$

Numerical Example …

- Applying a presymplectic Galerkin Hamiltonian variational integrator (with one internal stage) to this system yields a first-order method (with $m = g = L = 1$)

$$\begin{pmatrix} x_1 \\ (v_x)_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ (v_x)_0 \end{pmatrix} + \Delta t \begin{pmatrix} (v_x)_1 \\ \mathcal{P} x_1 \end{pmatrix},$$

$$\begin{pmatrix} (p_x)_1 \\ (p_{v_x})_1 \end{pmatrix} = \begin{pmatrix} (p_x)_0 \\ (p_{v_x})_0 \end{pmatrix} - \Delta t \left(\begin{pmatrix} 0 & 1 \\ \mathcal{P} & 0 \end{pmatrix}^T \begin{pmatrix} (p_x)_0 \\ (p_{v_x})_0 \end{pmatrix} + \begin{pmatrix} 2x_1 & 0 \\ (v_x)_1 & x_1 \\ 0 & 2(v_x)_1 \end{pmatrix}^T \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \\ \Lambda_3 \end{pmatrix} \right),$$

$$0 = \begin{pmatrix} x_1^2 + Y^2 - 1 \\ (v_x)_1 x_1 + V_y Y \\ (v_x)_1^2 + V_y^2 - Y + \mathcal{P} \end{pmatrix},$$

$$0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & x_1 \end{pmatrix}^T \begin{pmatrix} (p_x)_0 \\ (p_{v_x})_0 \end{pmatrix} + \begin{pmatrix} 2Y & 0 & 0 \\ V_y & Y & 0 \\ -1 & 2V_y & 1 \end{pmatrix}^T \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \\ \Lambda_3 \end{pmatrix}.$$

Numerical Example …

- For our numerical experiment, we apply the integrator to a collection of nearby initial positions $q_0 = (x_0, (v_x)_0)$ and a collection of nearby final momenta $p_1 = ((p_x)_1, (p_{v_x})_1)$
- Preservation of the presymplectic form $\Omega_0 = dx \wedge dp_x + dv_x \wedge dp_{v_x}$
  - Area occupied by the collection of points $(x_0, (p_x)_0)$ is the same as the area occupied by the collection of points $(x_1, (p_x)_1)$
  - Area occupied by the collection of points $((v_x)_0, (p_{v_x})_0)$ is the same as the area occupied by the collection of points $((v_x)_1, (p_{v_x})_1)$
- We compare the presymplectic method to the first-order method corresponding to using backward Euler in both position and momenta variables
- We take a large timestep $\Delta t = 2$ (roughly one-third of the period of the pendulum) to accentuate the difference between the two methods

Numerical Example …

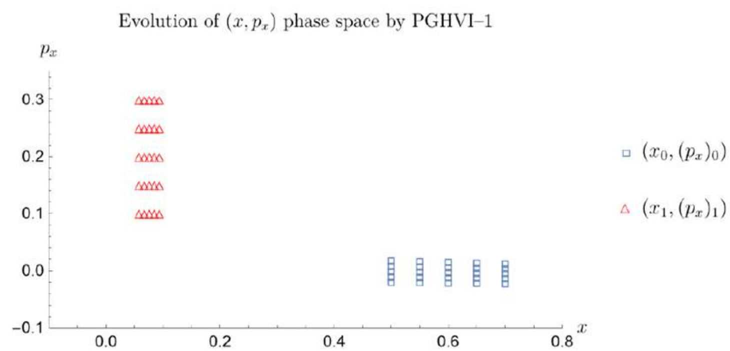


FIGURE 1. (x, p_x) phase space cross-section of PGHVI–1 applied to a distribution of initial conditions q_0 and final momenta p_1

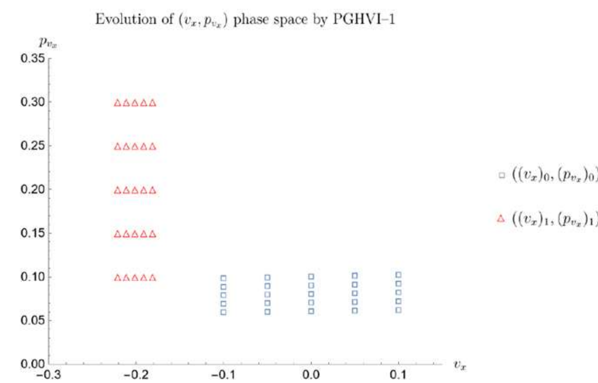


FIGURE 2. (v_x, p_{v_x}) phase space cross-section of PGHVI–1 applied to a distribution of initial conditions q_0 and final momenta p_1

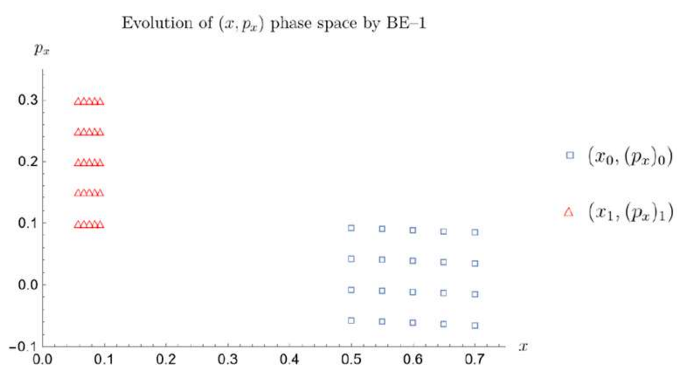


FIGURE 3. (x, p_x) phase space cross-section of BE–1 applied to a distribution of initial conditions q_0 and final momenta p_1

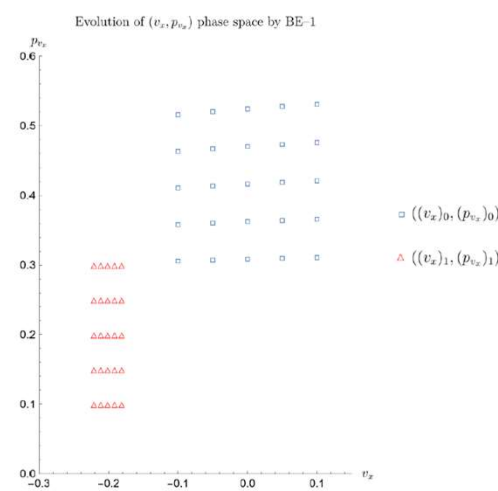


FIGURE 4. (v_x, p_{v_x}) phase space cross-section of BE–1 applied to a distribution of initial conditions q_0 and final momenta p_1

Numerical Example …

- Comparing the two methods for the Type II map $(q_0, p_1) \mapsto (q_1, p_0)$
  - The method using BE in both dynamical variables is implicit in position and explicit in momenta
  - The PGHVI method is implicit in position and implicit in momenta but linear in momenta (even when the DAE is nonlinear)
  - Both methods need to numerically solve the generally nonlinear constraint and adjoint constraint equations for the same number of algebraic variable internal stages
- The added cost for presymplecticity is solving a linear system in the momenta variable, even when the DAE is nonlinear. For high-dimensional problems, this of course can be significant; however, for example, in training a neural network, viewed as a discrete neural ODE, a bottleneck in training time is computing accurate gradients via backpropagation. Without (pre)symplecticity, have to use higher-order methods for accuracy.

Future Directions: Adjoint Systems for Evolution PDEs

- Example problem: PDE-constrained optimization

$$\min_u \Big[\frac{c_1}{2}\int_\Omega \Big(y(x,t_f) - \hat{y}(x,t_f)\Big)^2 dx + \frac{c_2}{2}\int_0^T \int_\Omega u(x,t)^2 dx dt\Big]$$
$$\text{s.t. } \partial_t y = \Delta y + f(y,u) \text{ on } (0,t_f)\times\Omega + \text{I.C.} + \text{B.C.}$$

  - Adjoint system arises as an extremization condition
- Questions:
  - Characterize geometry of adjoint systems for evolution PDEs?
    - e.g., infinite-dim. symplectic geometry, multisymplectic geometry
  - What can this geometry tell us about constructing geometric numerical methods for such problems?
    - Such problems require both temporal and spatial discretization
    - Natural choices for spatial semi-discretization?
    - Natural choice for temporal discretization?

Future Directions: Adjoint Systems for Evolution PDEs…

- Approach: extend theory to evolution PDE viewed as infinite-dim. ODE
- Abstract semilinear evolution equation:

X reflexive Banach space

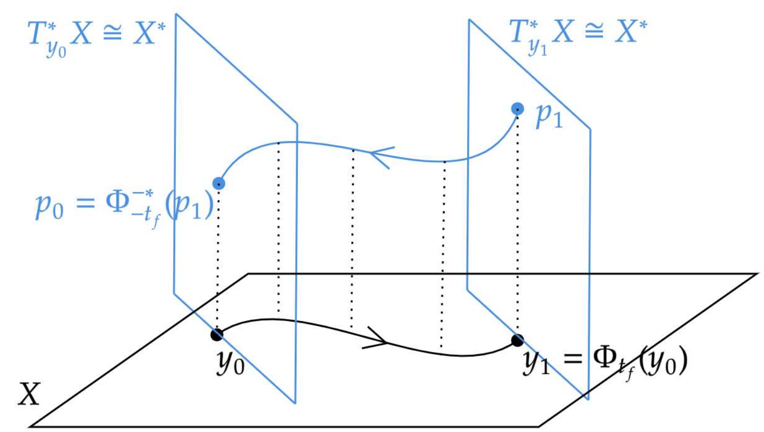
$$\dot{y} = Ay + f(y)$$
$$A : D(A) \subset X \to X$$
$$f : X \to X$$

- Define Hamiltonian and symplectic form on $T^*X \cong X \times X^*$

$$H : D(A) \times D(A^*) \to \mathbb{R}, \quad H(y,p) = \langle p, Ay + f(y) \rangle$$
$$\Omega(y,p) \cdot ((v_1, w_1), (v_2, w_2)) = \langle w_2, v_1 \rangle - \langle w_1, v_2 \rangle$$

- Adjoint system:

$$\dot{y} = Ay + f(y),$$
$$\dot{p} = -A^*p - [Df(y)]^*p$$



Φ (resp. Φ^{-*}) $\sim C^0$ semigroup generated by A (resp. $-A^*$)

Future Directions: Adjoint Systems for Evolution PDEs…

- Discretization:
  - Projection
    $$\Pi_h : X \to X_h \text{ (finite-dim. space)}$$
  - Semi-discrete symplectic form $\Omega_h = \Pi_h^{**}\Omega \in \Lambda^2(T^*X_h)$
  - Form associated semi-discrete adjoint ODE system
  - Integrate ODE system in time
- Many questions arise, e.g.,
  - Convergence? solution curves, symplectic form
  - Mild/weak solutions? Solutions with jumps (DG methods)?
  - Naturality: discretize then optimize versus optimize then discretize?
  - Constraints (PDAEs): infinite-dimensional presymplectic geometry?
  - Applications? e.g., optimization, neural PDEs

Some References

[1] BT and M. Leok. *Geometric Methods for Adjoint Systems.* Journal of Nonlinear Science, submitted, 2022.

[2] M. Leok and J. Zhang. *Discrete Hamiltonian variational integrators*. IMA J. Numer. Anal. 31(4):1497-1532, 2011.

[3] J. Schmitt and M. Leok. *Properties of Hamiltonian variational integrators*. IMA Journal of Numerical Analysis, 36(2), 2017.

[4] J. M. Sanz-Serna. *Symplectic Runge–Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more.* 58(1):3-33, 2016.

[5] BT and M. Leok. *Multisymplectic Hamiltonian variational integrators*. International Journal of Computer Mathematics (Special Issue on Geometric Numerical Integration, Twenty-Five Years Later), 99(1):113-157, 2022.

[6] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. *Neural Ordinary Differential Equations.* Advances in Neural Information Processing Systems, 31, 2018.