Geometric Adjoint Sensitivity Analysis for Lie Groups and PDEs

Melvin Leok

Mathematics, University of California, San Diego

Joint work with Ben Southworth and Brian Tran

AFOSR Dynamical Systems and Control Theory Review
BRICC, Arlington VA, August 2024

Funded by AFOSR FA9550-23-1-0279

mleok@ucsd.edu
http://mathweb.ucsd.edu/~mleok/

NSF DMS-1813635 • CCF-2112665 • DMS-2307801
AFOSR FA9550-18-1-0288 • FA9550-23-1-0279
DoD Newton Award • Simons Fellowship

Introduction and Motivation

Adjoint Systems

- Adjoint systems and their geometric discretization provide an efficient method of computing the parametric sensitivity of maps from a high-dimensional parameter space to a low-dimensional space.
- Such problems arise naturally in adjoint sensitivity analysis, adaptive mesh refinement, uncertainty quantification, automatic differentiation, superconvergent functional recovery, optimal control, optimal design, optimal estimation, and the training of neural ODEs.
- Most of these applications involve adjoint systems that arise from differential equations. We generalize to differential equations on Lie groups, differential-algebraic equations, and evolutionary PDEs, and how discretizations of the adjoint system can be constructed so that discretization commutes with forming adjoints.

Sensitivity Analysis

Problem Statement

- **Sensitivity analysis** determines how an objective function, such as a terminal or running cost, subject to the flow of an ODE or DAE, with respect to a perturbation in the initial condition.
- Given a terminal cost $C(q(t_f))$, and an ODE $\dot{q} = f(q)$ on a manifold M with initial condition $q(0) = q_0$, the sensitivity question is how does the cost change infinitesimally if the initial condition is perturbed by δq_0?

Sensitivity Analysis

Direct Method

- The **direct method** involves the **variational equation**,

$$\dot{q} = f(q), \quad \dot{v} = Df(q)v,$$

which is an ODE on TM corresponding to the tangent lift of f.

- The **sensitivity** induced by δq_0 is

$$\delta C(t_f) = \langle \nabla_q C(q(t_f)), v(t_f) \rangle,$$

where we have chosen the initial conditions $q(0) = q_0$, $v(0) = \delta q_0$.

- For many optimal control problems, we are interested in finding the δq_0 that induces a desired sensitivity $\delta C(t_f)$.
- Computing this involves $\mathcal{O}(N)$ integrations of the variational equations, where N is the number of design parameters. This is prohibitive when N is large.

Sensitivity Analysis

Adjoint Method

- The **adjoint system** is given by
$$\dot{q} = f(q), \quad \dot{p} = -[Df(q)]^* p,$$
which is the cotangent lift of $f \in \mathfrak{X}(M)$ to $\hat{f} \in \mathfrak{X}(T^*M)$.
- The solution curves (q,p) of the adjoint system and (q,v) of the variational system that cover the same base curve q satisfy an adjoint-variational **quadratic conservation law**,
$$\frac{d}{dt}\langle p, v\rangle = 0.$$
- Given Type II boundary conditions $q(0) = q_0$, $p(t_f) = \nabla_q C(q(t_f))$, the quadratic conservation law implies that $\delta C(t_f) = \langle p(0), \delta q_0\rangle$.
- This requires $\mathcal{O}(N_C)$ integrations, where N_C is the number of cost functions. The adjoint methods is advantageous when $N_C \gg N$.

Geometric Characterization of Adjoint Systems

Formal Hamiltonians and Adjoint Equations

- The adjoint system can be viewed as a Hamiltonian system on T^*M, in terms of a **formal Hamiltonian** $H : T^*M \to \mathbb{R}$,
$$H(q,p) = \langle p, f(q) \rangle.$$
- The adjoint system is precisely the Hamilton's equations for the formal Hamiltonian with respect to $\Omega = dq \wedge dp$, i.e., $i_{\hat{f}}\Omega = dH$.
- The Hamiltonian flow is symplectic, and symplecticity implies the quadratic conservation law. Symplecticity states that along a solution curve of the adjoint system, $\frac{d}{dt}\omega(V,W) = 0$, where V and W are first variations to the adjoint system. For first variations $V = v\,\partial/\partial q$ and $W = p\,\partial/\partial p$, we have that $\omega(V,W) = \langle p, v \rangle$, which implies that $\langle p, v \rangle$ is preserved.

Geometric Characterization of Adjoint Systems

Type II Variational Principle

- Type I boundary conditions do not make sense for adjoint systems. In general, Type II boundary conditions do not make intrinsic sense, as one cannot specify a covector without specifying a basepoint. But since adjoint systems cover an ODE, given q_0, integrating the ODE gives $q_1 = \Phi_{t_f}(q_0)$.
- The intrinsic Type II variational principle is given by,

$$\delta \int_0^{t_f} [\langle p, \dot{q} \rangle - H(q,p)] = \delta \int_0^{t_f} \langle p, \dot{q} - f(q) \rangle dt = p(t_f)\delta q(t_f),$$

with Type II boundary conditions,

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

Geometric Characterization of Adjoint Systems

Continuous Adjoint Systems for DAEs

- Consider a DAE,

$$\dot{q} = f(q, u), \qquad 0 = \phi(q, u),$$

with dynamic variables q and algebraic variables u.

- The formal Hamiltonian is given by,

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

and the adjoint system is,

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} = f(q, u), & \dot{p} &= -\frac{\partial H}{\partial q} = -[D_q f(q, u)]^* p - [D_q \phi(q, u)]^* \lambda, \\ 0 &= \frac{\partial H}{\partial \lambda} = \phi(q, u), & 0 &= -\frac{\partial H}{\partial u} = -[D_u f(q, u)]^* p - [D_u \phi(q, u)]^* \lambda. \end{aligned}$$

Geometric Discretization of Adjoint Systems

Geometric Integration of Adjoint Systems for ODEs

- Sanz-Serna showed that discretization commutes with adjoints is the adjoint system is integrated with a symplectic Runge–Kutta method that covers the original Runge–Kutta method.
- This discretization preserves a discrete quadratic conservation law, and the following diagram commutes,

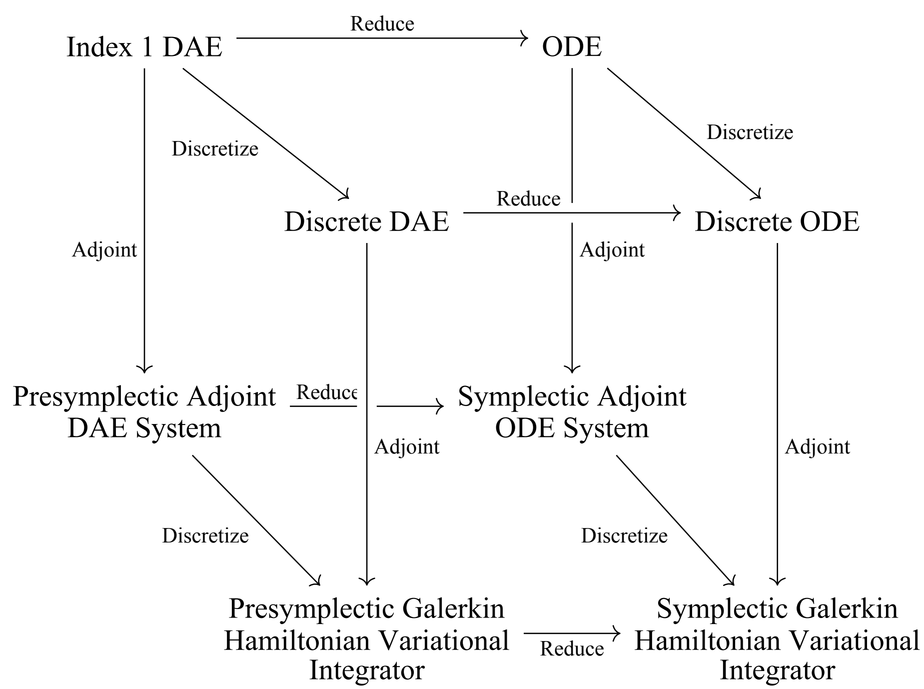
$$\begin{tikzcd}
\text{ODE} \arrow[r, "\text{Discretize}"] \arrow[d, "\text{Adjoint}"'] & \begin{array}{c}\text{Runge–Kutta}\\ \text{discretization}\end{array} \arrow[d, "\text{Adjoint}"] \\
\text{Adjoint System} \arrow[r, "\text{Discretize}"'] & \begin{array}{c}\text{Symplectic}\\ \text{partitioned RK}\end{array}
\end{tikzcd}$$

- Analogous to the Ross–Fahroo lemma in optimal control, where dualization and discretization commutes for covector mappings.

Geometric Discretization of Adjoint Systems

Geometric Integration of Adjoint Systems for DAEs

- We extended this to index-one DAEs using presymplectic geometry, and showed the index is the iterations necessary for the Gotay–Nester–Hinds algorithm[1] to reduce the presymplectic system.



[1]an alternating projection method for finding $P \subset T^*Q$, such that: (i) Hamilton's equations are consistent; (ii) they define a vector field tangent to P

Geometric Discretization of Adjoint Systems

Variational Discretization of Lagrangian Systems

- **Discrete Lagrangian**

$$L_d(q_0, q_1) \approx L_d^{\text{exact}}(q_0, q_1) \equiv \underset{\substack{q \in C^2([0,h],Q) \\ q(0)=q_0, q(h)=q_1}}{\text{ext}} \int_0^h L(q(t), \dot{q}(t))dt.$$

- **Discrete Hamilton's principle**

$$\delta \mathbb{S}_d = \delta \sum L_d(q_k, q_{k+1}) = 0,$$

where q_0, q_N are fixed.

- **Discrete Euler-Lagrange equation**

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) = 0,$$

which is equivalent to,

$$p_k = -D_1 L_d(q_k, q_{k+1}), \quad p_{k+1} = D_2 L_d(q_k, q_{k+1}).$$

Geometric Discretization of Adjoint Systems

Variational Discretization of Hamiltonian Systems

- A Hamiltonian $H : T^*Q \to \mathbb{R}$ is **degenerate** if the **Legendre transformation** $\mathbb{F}H : T^*Q \to TQ$, $(q, p) \mapsto (q, \partial H/\partial p)$, is non-invertible. The formal Hamiltonian is degenerate.
- This obstructs the construction of variational integrators for degenerate Hamiltonian systems by traversing via the Lagrangian side.

$$\begin{tikzcd} H(q,p) \arrow[r, "\mathbb{F}H"] \arrow[d, dashed] & L(q,\dot{q}) \arrow[d] \\ H_d^+(q_0,p_1) & L_d(q_0,q_1) \arrow[l, "\mathbb{F}L_d"'] \end{tikzcd}$$

- The goal is to **construct discrete Hamiltonians directly**, so that the diagram commutes for hyperregular Hamiltonians.

Geometric Discretization of Adjoint Systems

Exact Discrete Hamiltonian

- The exact discrete Lagrangian is a Type I generating function,

$$L_d^{\text{exact}}(q_0, q_1) \equiv \operatorname*{ext}_{\substack{q \in C^2([0,h],Q) \\ q(0)=q_0, q(h)=q_1}} \int_0^h L(q(t), \dot{q}(t)) dt.$$

- Continuous Legendre transform to obtain $L(q, \dot{q}) = p\dot{q} - H(q, p)$.
- Discrete Legendre transform for a Type II generating function,

$$H_{d,\text{exact}}^+(q_k, p_{k+1}) =$$

$$\operatorname*{ext}_{\substack{(q,p) \in C^2([t_k, t_{k+1}], T^*Q) \\ q(t_k)=q_k, p(t_{k+1})=p_{k+1}}} p(t_{k+1})q(t_{k+1}) - \int_{t_k}^{t_{k+1}} \left[p\dot{q} - H(q,p)\right] dt.$$

- Choose an approximation space for Q (not T^*Q) and quadrature.

Adjoint Systems on Lie Groups

Adjoint Systems on Lie Groups I

- Many optimization and optimal control problem occur on Lie groups.
- Given an ODE $\dot{g} = F(g)$ on a Lie group G, the adjoint system is an ODE on T^*G, and we can make sense of the Type II variational principle globally via left or right trivialization.
- With respect to a left-trivialization of T^*G, $(g, p) \mapsto (g, \mu) = (g, g^* p) \in G \times \mathfrak{g}^*$, the adjoint system has the form,

$$\dot{g} = F(g), \quad \dot{\mu} = -g^* \cdot [Df(g)]^* \mu + \mathrm{ad}^*_{f(g)} \mu,$$

  where $f(g) = g^{-1} F(g)$.
- More generally, we developed a Type II variational principle for any Hamiltonian system on T^*G.

Adjoint Systems on Lie Groups

Adjoint Systems on Lie Groups II

- The left-trivialized formal Hamiltonian is $h(g,\mu) = \langle \mu, f(g)\rangle$.
- The Lie–Poisson equations hold on $G \times \mathfrak{g}^*$,

$$\begin{aligned}\dot{g} &= g \cdot D_\mu h(g,\mu),\\ \dot{\mu} &= -g^* \cdot D_g h(g,\mu) + \mathrm{ad}^*_{D_\mu h(g,\mu)}\mu,\end{aligned}$$

with Type II boundary conditions, $g(0) = g_0$, $\mu(t_f) = \mu_1$.

Geometric Discretization

- The discrete Lie–Poisson adjoint equations are,

$$\begin{aligned}(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m_k &= -\Delta t g_k^*[Df(g_k)]^* m_{k+1},\\ \xi_{k+1} &= f(g_k),\\ g_{k+1} &= g_k\tau(\Delta t\xi_{k+1}).\end{aligned}$$

Adjoint Systems for Evolutionary PDEs

Adjoint Systems for Evolutionary PDEs

- Consider a semilinear evolution equation of the form,

$$\dot{q}(t) = Aq(t) + f(t, q(t)), \quad q(0) = q_0.$$

- To view the adjoint system of this as an infinite-dimensional Hamiltonian system, we consider time-dependent Hamiltonian systems,

$$i_{X_H}(\Omega - dH \wedge dt) = 0,$$

where Ω is the pullback of the spatial symplectic form to spacetime.

- The adjoint Hamiltonian is $H(t, q, p) = \langle p, Aq + f(t, q)\rangle$.
- The adjoint system is

$$\dot{q} = Aq + f(t, q), \quad \dot{p} = -A^* p - [Df(t, q)]^* p.$$

- As before, there is an adjoint-variational quadratic invariant.

Adjoint Systems for Evolutionary PDEs

Spatial Semidiscretization

- $\{\varphi_i\}_{i=1}^{\dim(X_h)}$ is a basis for X_h, and $\{l_j\}_{j=1}^{\dim(X_h)}$ is a basis for X_h^*.
- A **Galerkin semidiscretization** is specified by an approximation $q(t) \approx \sum_i \mathbf{q}^i(t)\varphi_i$ satisfying,

  $\left\langle l_j, \dot{\mathbf{q}}^i(t)\varphi_i - \mathbf{q}^i(t)A\varphi_i - f(t, \mathbf{q}^k(t)\varphi_k)\right\rangle = 0,\ j = 1, \dots, \dim(X_h),$
- Let M and K denote mass and stiffness matrices,

  $$M_{ji} = \langle l_j, \varphi_i\rangle, \quad K_{ji} = \langle l_j, A\varphi_i\rangle,$$

  and the semidiscretized semilinear term is

  $$\mathbf{f}^j(t, \mathbf{q}) = \langle l_j, f(t, \mathbf{q}^k\varphi_k)\rangle.$$

  Then,

  $$M\frac{d}{dt}\mathbf{q} = K\mathbf{q} + \mathbf{f}(t, \mathbf{q}).$$

Adjoint Systems for Evolutionary PDEs

Naturality of the Full Discretization

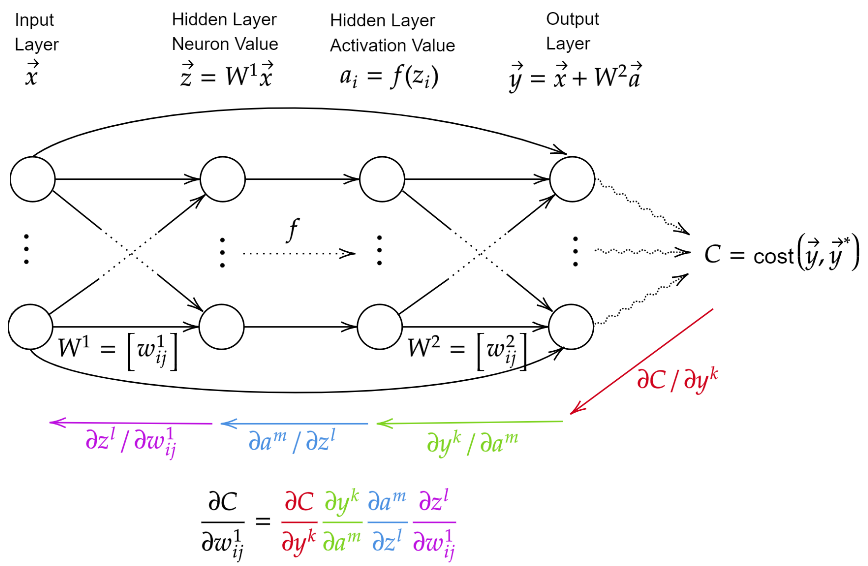
- The continuous, semidiscretization and full discretization, and their quadratic conservation laws, are related as follows,

$$
\begin{tikzcd}
\dot q = g \arrow[rr, "\text{Adjoint}"] \arrow[d, "S"] & & \begin{array}{c} \dot q = g \\ \dot p = -[D_q g]^* p \end{array} \arrow[r, dash, dashed] \arrow[d, "{(S,S^*)}"] & \frac{d}{dt}\langle p, \delta q\rangle = 0 \\
\dot{\mathbf{q}} = M^{-1}\mathbf{g} \arrow[r, "\text{Adjoint}_S"] \arrow[rr, bend left, "\text{Adjoint}_M"] \arrow[d, "\Phi_{\Delta t}"] & \begin{array}{c} \dot{\mathbf{q}} = M^{-1}\mathbf{g} \\ \dot{\mathbf{z}} = -[D_{\mathbf{q}}\mathbf{g}]^T M^{-T}\mathbf{z} \end{array} \arrow[r, shift left, "\mathbf{z}=M^T\mathbf{p}"] \arrow[d, "{(\Phi_{\Delta t}, T^{*S}\Phi^{-1}_{\Delta t})}"] & \begin{array}{c} \dot{\mathbf{q}} = M^{-1}\mathbf{g} \\ \dot{\mathbf{p}} = -M^{-T}[D_{\mathbf{q}}\mathbf{g}]^T \mathbf{p} \end{array} \arrow[l, shift left] \arrow[r, dash, dashed] \arrow[d, "{(\Phi_{\Delta t}, T^{*M}\Phi^{-1}_{\Delta t})}"] & \frac{d}{dt}\langle \mathbf{p}, \delta \mathbf{q}\rangle_M = 0 \\
\mathbf{q}_{n+1} = \Phi_{\Delta t}(\mathbf{q}_n) \arrow[r, "\text{Adjoint}_S"] \arrow[rr, bend right, "\text{Adjoint}_M"'] & \begin{array}{c} \mathbf{q}_{n+1} = \Phi_{\Delta t}(\mathbf{q}_n) \\ \mathbf{z}_{n+1} = T^{*S}\Phi^{-1}_{\Delta t}(\mathbf{z}_n) \end{array} \arrow[r, dashed, shift left, "\text{similarity}"] & \begin{array}{c} \mathbf{q}_{n+1} = \Phi_{\Delta t}(\mathbf{q}_n) \\ \mathbf{p}_{n+1} = T^{*M}\Phi^{-1}_{\Delta t}(\mathbf{p}_n) \end{array} \arrow[l, dashed, shift left] \arrow[r, dash, dashed] & \begin{array}{c} \langle \mathbf{p}_{n+1}, \delta \mathbf{q}_{n+1}\rangle_M \\ = \langle \mathbf{p}_n, \delta \mathbf{q}_n\rangle_M \end{array}
\end{tikzcd}
$$

Application of Geometric Adjoint Sensitivity

Training of Neural ODEs

- Residual and recurrent neural networks can be modeled as
$$x_{t+1} = x_t + g(t, x_t, W(t)),$$
and viewed as a discretization of an ODE, $\dot{x} = g(t, x(t), W(t))$.
- Symplectic discretization of the adjoint system of the neural ODE is an alternative to backpropagation using automatic differentiation.



Summary and Future Directions

- Adjoint systems have a Hamiltonian structure and Hamiltonian variational integrators allow the direct and indirect approach to be equivalent for DAEs and evolutionary PDEs. The geometric discretizations of the adjoint system can be constructed using Hamiltonian variational integrators.
- Such methods can be used to train physics-informed neural networks, and neural PDEs. Many novel structured neural network architectures can be viewed as the discretization of an ODE with geometric invariants, for which these methods are also applicable.
- It would be interested to study these properties in terms of Dirac mechanics and geometry, and to consider the generalization to multi-Dirac field theories and geometry for the geometric adjoint sensitivity analysis of PDEs.

References

1. J.M. Sanz-Serna, *Symplectic Runge–Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more*, SIAM Review, 51(1), 3–33, 2016.
2. B.K. Tran, M. Leok, *Geometric methods for adjoint systems*, Journal of Nonlinear Science, 34:25 (75 pages), 2024.
3. B.K. Tran, M. Leok, *Type II Hamiltonian Lie group variational integrators with applications to geometric adjoint sensitivity analysis*, arXiv:2311.03527
4. B.K. Tran, B.S. Southworth, M. Leok, *On properties of adjoint systems for evolutionary PDEs*, Journal of Nonlinear Science, 34:95 (39 pages), 2024.
5. M. Leok, J. Zhang, *Discrete Hamiltonian variational integrators*, IMA Journal of Numerical Analysis, 31(4), 1497–1532, 2011.