Geometric Methods for Adjoint Systems

Brian Tran (joint work with Prof. Melvin Leok)
Department of Mathematics, University of California, San Diego

Motivation

- Adjoint Systems are used to efficiently compute the sensitivity of a terminal or running cost function
$$C(q(t_f)) \text{ or } \int_0^{t_f} L(q(t))dt$$
subject to an ordinary differential equation (ODE) constraint
$$\dot{q}(t) = f(q(t)),\ q(0) = q_0,$$
with respect to a perturbation in the initial condition δx_0.
- Adjoint systems arise as the extremization conditions for optimal control problems via the Pontryagin maximum principle.

Hamiltonian Description of Adjoint Systems

- Consider an ODE $\dot{q} = f(q)$, specified by a vector field on a manifold M, $f \in \Gamma(TM)$.
- Define the adjoint Hamiltonian $H : T^*M \to \mathbb{R}$ by
$$H(q,p) = \langle p, f(q)\rangle.$$
- The adjoint system is given by a Hamiltonian system on T^*M relative to the canonical symplectic form $\Omega = dq \wedge dp$,
$$i_{X_H}\Omega = dH.$$
- In coordinates, an integral curve of X_H has the expression
$$\dot{q} = \partial H/\partial p = f(q),$$
$$\dot{p} = -\partial H/\partial q = -[Df(q)]^* p.$$
- The Hamiltonian vector field X_H is the cotangent lift of f to a vector field on T^*M.

Symplecticity and Adjoint Sensitivity Analysis

- Since the adjoint system is Hamiltonian, the flow of the system is symplectic; i.e., it preserves the symplectic form Ω. This can be expressed
$$\frac{d}{dt}\Omega_{(q(t),p(t))}(V(t), W(t)) = 0,$$
where V and W are first variations of the adjoint system, which can be identified with solutions of the linearization of the adjoint system
$$\frac{d}{dt}\delta q = Df(q)\delta q,$$
$$\frac{d}{dt}\delta p = -[Df(q)]^*\delta p.$$
- Symplecticity implies the quadratic conservation law
$$\frac{d}{dt}\langle p(t), \delta q(t)\rangle = 0.$$
- Adjoint Sensitivity Analysis: By the above, $\langle p(t_f), \delta q(t_f)\rangle = \langle p(0), \delta q(0)\rangle$. Choosing $p(t_f) = \nabla_q C(q(t_f))$, one can backpropagate to solve for $p(0)$, which, by the quadratic conservation law, gives the sensitivity of a terminal cost function with respect to a perturbation in the initial condition
$$p(0) = \frac{\partial}{\partial \delta q_0} C(q(t_f)).$$
- Can similarly treat a running cost function, by augmenting the Hamiltonian $H_L(q,p) = H(q,p) + L(q)$.

Differential-Algebraic Equations

- Consider a differential-algebraic equation (DAE)
$$\dot{q} = f(q,u),$$
$$0 = \phi(q,u).$$
Here, $q \in M_d$ are the dynamical variables and $u \in M_a$ are the algebraic variables. Geometrically, a DAE is specified by a section f of the bundle $\overline{TM_d}$, the pullback bundle of TM_d by $M_d \times M_a \to M_d$, and by a section ϕ of a vector bundle $\Phi \to M_d \times M_a$.
- Say that the DAE has index 1 if $\partial\phi/\partial u$ is an isomorphism pointwise. By the implicit function theorem, one can locally solve the constraint equation for $u = u(q)$ and reduce the DAE to an ODE
$$\dot{q} = f(q, u(q)).$$

Adjoint Systems for DAEs

- Idea: extend the notion of an adjoint system to DAEs.
- To do this, introduce the spaces

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

- Define the adjoint DAE Hamiltonian $H : \overline{T^*M_d} \oplus \Phi^* \to \mathbb{R}$ by
$$H(q,u,p,\lambda) = \langle p, f(q,u)\rangle + \langle \lambda, \phi(q,u)\rangle.$$
- Using the above maps, pullback the symplectic form Ω on T^*M_d to a presymplectic form Ω_0 on $\overline{T^*M_d} \oplus \Phi^*$.
- Define the adjoint DAE system as the presymplectic Hamiltonian system
$$i_{X_H}\Omega_0 = dH.$$
- In coordinates,
$$\dot{q} = \partial H/\partial p = f(q,u),$$
$$\dot{p} = -\partial H/\partial q = -[D_q f(q,u)]^* p - [D_q \phi(q,u)]^*\lambda,$$
$$0 = \partial H/\partial \lambda = \phi(q,u),$$
$$0 = -\partial H/\partial u = -[D_u f(q,u)]^* p - [D_u \phi(q,u)]^*\lambda.$$
- The vector field X_H is in general only defined on the primary constraint submanifold specified by the last two equations. However, the flow of X_H may leave the submanifold, so one must further restrict to a final constraint submanifold to which X_H is tangent. This process to obtain such a final constraint submanifold is known as the presymplectic constraint algorithm.
- When the underlying DAE has index 1, the presymplectic constraint algorithm terminates after one step; i.e., the primary and final constraint submanifolds coincide.
- Presymplecticity of the flow of X_H yields a quadratic conservation law analogous to the ODE case, allowing one to compute sensitivities of a terminal or running cost function subject to a DAE constraint.

Structure-Preserving Discretizations of Adjoint Systems

- In most cases, one cannot analytically solve an adjoint system; hence, one must discretize the system; i.e., numerically integrate the system.
- Key Idea: since an adjoint system has a (pre)symplectic structure, it is natural to utilize a (pre)symplectic integrator to discretize such systems. In particular, such integrators preserve the (pre)symplectic form and hence, preserve the quadratic conservation laws used for adjoint sensitivity analysis.
- We study how Galerkin Hamiltonian variational integrators can be used to discretize such systems and extend the construction of these integrators to presymplectic systems.
- We show that the process of forming an adjoint system, discretizing, and reducing (from an index 1 DAE to an ODE) commute, for particular choices of these processes:

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

- Using this naturality, we show that if the discrete generating function approximates the exact generating function to order r, then the Type II flow $(q_0, p_1) \mapsto (q_1, p_0)$ map is order-r accurate.

Future Research Direction

- We aim to explore the extension of this framework to the setting of infinite-dimensional PDEs; in particular, to develop geometric methods for adjoint systems for semilinear evolution equations
$$\dot{q} = Aq + f(q),$$
where A is an unbounded operator on a Banach space and f is a nonlinear operator on a Banach space.
- The main tools are infinite-dimensional symplectic geometry and the theory of C_0-semigroups. For discretization, we will utilize the Galerkin method in space and symplectic integration in time, with the aim of proving an extended naturality result.

Summary

- The utility of adjoint systems for computing sensitivies can be understood through (pre)symplectic geometry.
- One can utilize geometric integration to preserve the structures relevant to adjoint sensitivity analysis and hence, construct integrators which can be used to exactly compute sensitivities.

Supported by NSF Graduate Research Fellowship DGE-2038238, NSF Grants DMS-1813635, DMS-1345013, AFOSR Grant FA9550-18-1-0288, DoD Newton Award for Transformative Ideas, Simons Fellowship in Mathematics