Topology of Optimal Control Landscapes for Classical Mechanical Systems

Carlee Joe-Wong

September 16, 2010

1 Introduction

Optimal control of quantum dynamics phenomena has emerged as an important research field in the last twenty years. Generally, such control is achieved by designing laser pulses to maximize the probability of an event (e.g., plasma ionization) occurring after a chosen time period. The system is assumed to follow quantum dynamics; subject to this constraint, a laser field is designed that minimizes or maximizes a cost functional aiming to reach a target state. This field is usually found through an iterative process of solving the system's Euler-Lagrange equations, starting from an initial guess and employing either probabilistic (e.g., simulated annealing) or deterministic (e.g., gradient search) methods to evolve towards an optimal control field.

The cost functional-dynamic constraint formulation of controlling quantum phenomena like molecular dynamics takes its inspiration from optimal control theory, an engineering discipline; the principal difference is that in engineering applications, systems are assumed to follow classical mechanics. Yet molecular systems are often well-described by classical mechanics; if the molecular system is taken to follow classical instead of quantum dynamics, with a classically-specified result, the control problem often becomes more numerically tractable, especially for polyatomic systems. Results from some papers, e.g. [8], suggest that classical and quantum control methods should give similar results.

It has recently been shown in [5], [6] that the quantum control landscape is especially "nice" in that for a certain type of cost functional, there are no traps, or local extrema where a gradient algorithm might get stuck. This mathematical result followed several papers demonstrating much greater success in quantum control than might be expected, given the problem's large dimensionality. In this project, I worked with Professor Herschel Rabitz and Tak-San Ho to extend analysis of the quantum control topological landscape to the classical control one.

2 Numerical Examples

Throughout this project, I focus on cost functionals of the form

$$J = \Phi(x(T)), \tag{1}$$

where $x = [q, p]^T$ is the system state and T is the fixed final time of the simulation. The variables q and p represent position and momentum respectively; there are n of each, where n is the number of system degrees of freedom. The system dynamics are assumed to be

$$\dot{q} = \frac{\partial H}{\partial p} \tag{2}$$

$$\dot{p} = -\frac{\partial H}{\partial q}, \tag{3}$$

where

$$H(t) = \frac{1}{2} p^T G p + V(q) - D(q)\epsilon(t) \tag{4}$$

is the system's energy Hamiltonian. Here G is the Wilson G-matrix, V is the potential energy function, D is the dipole moment function and ϵ is the system's control field, assumed to be aligned with D. For simplicity, let $f = \left[\frac{\partial H}{\partial p} \; -\frac{\partial H}{\partial q}\right]^T$. I analyze the optimization problem

$$\min J = \Phi(x(T)) \tag{5}$$

$$\text{s.t. } \dot{x} = f(x, u) \tag{6}$$

$$\text{var. } \epsilon(t), t \in [0, T]$$

I ran numerical simulations for three types of systems–linear diatomic, linear triatomic, and chaotic, with quartic potential. The results all indicated an absence of numerical traps, though in some cases pushing the cost to 0 required prohibitive computational time. The algorithm for computing an optimal field is outlined in Algorithm 1. The s parameter measures the simulation time; it

Algorithm 1 Gradient algorithm to find an optimal control field.

1: Choose a random initial field ϵ.
2: Compute $\frac{\delta J}{\delta \epsilon}$ at discrete time points $\delta t, 2\delta t, \ldots, T$.
3: Let $\frac{d\epsilon}{ds} = -\frac{\delta J}{\delta \epsilon}$. Evolve ϵ accordingly using the fourth order Runge-Kutta algorithm.

generalizes the iteration number. The standard Euler-Lagrange equations

$$\lambda(T) = \frac{\partial J}{\partial x} \tag{7}$$

$$\dot{\lambda}^T = -\lambda^T \frac{\partial f}{\partial x} \tag{8}$$

$$\frac{\delta J}{\delta \epsilon(t)} = \lambda^T \frac{\partial f}{\partial \epsilon(t)}, \tag{9}$$

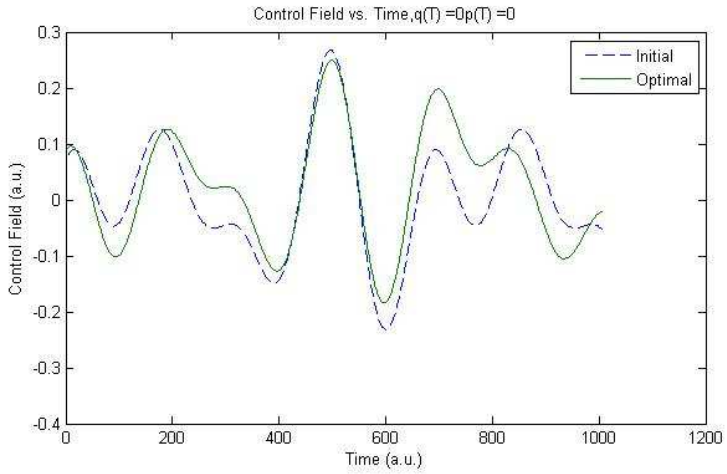

Figure 1: Control field evolution, target and final (q, p) = (0,0) and (1.728e-5,-4.50e-6).

as derived in [7], were used to compute $\frac{\delta J}{\delta \epsilon}$. Here λ represents the time-varying Lagrange multiplier. With this algorithm, the cost decreases throughout the simulation; there is no indication that it ever gets stuck before cost goes to 0.

Figure 1 shows the initial and final control fields for Hamiltonian

$$H(t) = \frac{p^2}{2m} + D\left(1 - e^{-\alpha q}\right)^2 - Aqe^{-\xi q^4}\epsilon, \tag{10}$$

where $m = 1732$, $D = 0.2101$, $A = 0.4541$ and $\xi = 0.0064$; this Hamiltonian models an HF molecule. The final time was 320π a.u., with time-step $\delta t = \frac{10\pi}{9}$ a.u. The cost functional was

$$J = p^2 + q^2, \tag{11}$$

so that the target state was $(q, p) = (0, 0)$. These results, where the cost was driven almost to 0, show a typical optimal system. Figure 2 shows how the cost evolves over the control field evolution. As can be seen, numerical errors cause cost to oscillate once it gets very small, but the overall trend is still for the cost to decrease to 0. Figure 3 shows the final position and momentum trajectories; these oscillation patterns are typical of other simulations. Indeed, the initial control field probably evolves relatively little is probably because a small adjustment of the trajectory will lead to a phase shift of the q and p trajectories, enabling them to hit their target values.

3 Non-singular (Regular) Controls

I define regular controls as those $\epsilon(t)$ such that the mapping $\epsilon \to x(T)$, where as above $x(T)$ represents the final state of the system, is surjective. This means

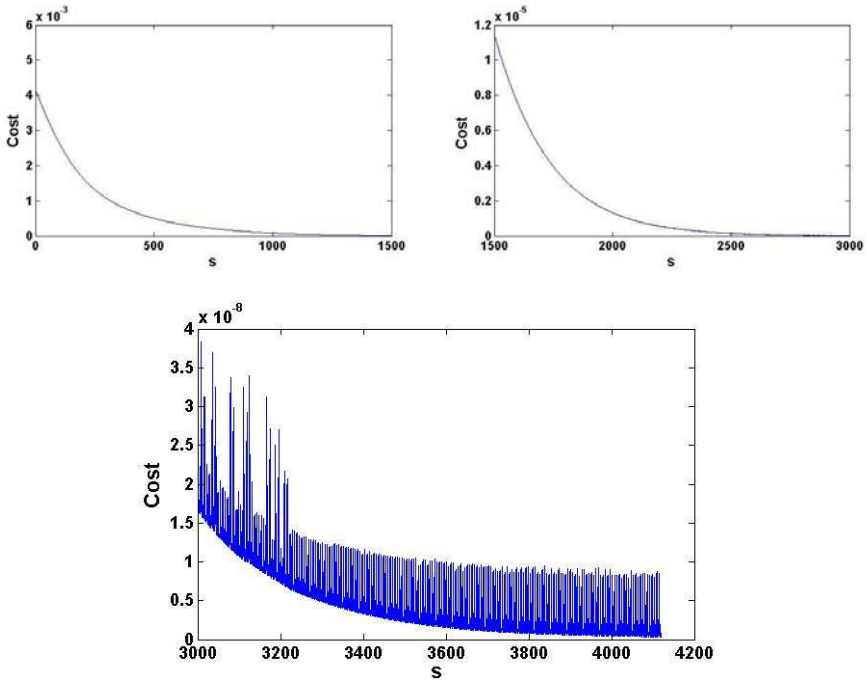

Figure 2: Cost evolution, target and final (q, p) = (0,0) and (1.728e-5,-4.50e-6).

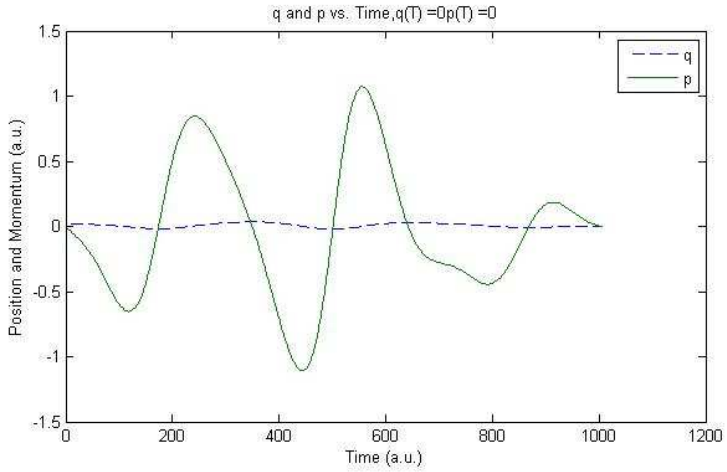

Figure 3: Final position and momentum trajectories, target and final (q, p) = (0,0) and (1.728e-5,-4.50e-6).

that the $2n \times 1$ Fréchet derivative $\frac{\delta x(T)}{\delta \epsilon(t)}$ must consist of $2n$ linearly independent functions, where n is the number of degrees of freedom in the system.

The gradient algorithm attempts to drive the gradient $\frac{\delta J}{\delta \epsilon}$, as computed in Algorithm 1, to 0. Thus, to find out if the gradient algorithm will ever get trapped, I need to characterize those points at which $\frac{\delta J}{\delta \epsilon} = 0$. Using the chain rule,

$$\frac{\delta J}{\delta \epsilon(t)} = \frac{\partial J}{\partial x(T)} \frac{\delta x(T)}{\delta \epsilon(t)}, \tag{12}$$

which must be 0 for all $t \in [0, T]$. Given that this control ϵ is nonsingular, $\frac{\delta x(T)}{\delta \epsilon(t)}$ has rank $2n$, where n is the number of degrees of freedom. Thus, there is no nonzero vector v satisfying $v \frac{\delta x(T)}{\delta \epsilon(t)} = 0$ for all t, and $\frac{\delta J}{\delta \epsilon(t)} = 0$ implies $\frac{\partial J}{\partial x(T)} = 0$.

The problem of characterizing the dynamic system's critical points has now become the kinematic problem of finding the critical points of J, as a function solely of $x(T)$, the system's final state. To do this, choose

$$J = \big(O(x(T)) - O_t\big)^T \big(O(x(T)) - O_t\big), \tag{13}$$

where O is a vector-valued system observable with length r and target value O_t. Then

$$\frac{\partial J}{\partial x(T)} = 2\big(O(x(T)) - O_t\big)^T \frac{\partial O}{\partial x(T)}. \tag{14}$$

If $O(x(T)) - O_t = 0$, then $J = 0$ and the control gives a global optimum. However, if $O(x(T)) - O_t \neq 0$ but $\frac{\partial O}{\partial x(T)} = 0$, then the corresponding control is a critical point that is not a global optimum. This situation can be avoided if O has no local extrema in $x(T)$, possibly reaching a global extremum only at the target O_t. Even if O does have other extrema, however, this critical point may not be a non-global minimum. That only occurs if the Hessian $\frac{\delta^2 J}{\delta \epsilon(t)^2}$ is positive definite, positive semidefinite or negative semidefinite (with the latter two cases requiring higher order derivative tests).

Let us assume $\frac{\partial O}{\partial z(T)} \neq 0$. If O is scalar, clearly $\frac{\partial J}{\partial x(T)} = 0$ implies $O(x(T)) = O_t$. However, if O is not scalar, then it is possible for $\big(O(x(T)) - O_t\big) \neq 0$ and $\frac{\partial O}{\partial x(T)} \neq 0$, while $\frac{\partial J}{\partial x(T)} = 0$. However, this does not happen if the $r \times 2n$ Jacobian matrix $\frac{\partial O}{\partial x(T)}$ has rank r. Letting $A = \frac{\partial O}{\partial x(T)}$, for that case (14) gives

$$2\big(O(x(T)) - O_t\big)^T AA^T (AA^T)^{-1} = (0) A^T (AA^T)^{-1} = 0, \tag{15}$$

where AA^T is invertible because it is an $r \times r$ matrix of rank r. Then $O(x(T)) - O_t = 0$ and this control field is globally optimal. For instance, if O_t is a target state, $\left(\frac{\partial O}{\partial x(T)}\right)^T$ is a full-rank matrix of 0's and 1's, with rank equal to r, the number of state variables being observed.

Now suppose that the Jacobian has rank $k < r$. This is always the case if $r > 2n$, since then $\operatorname{rank}\left(\frac{\partial O}{\partial x(T)}\right) \leq 2n < r$. We know that $\big(O(x(T)) - O_t\big)^T \frac{\partial O}{\partial x(T)} = 0$

is equivalent to

$$\left(\left(O(x(T)) - O_t\right)^T \frac{\partial O}{\partial x(T)}\right)^T = \left(\frac{\partial O}{\partial x(T)}\right)^T \left(O(x(T)) - O_t\right) = 0, \qquad (16)$$

where $\left(\frac{\partial O}{\partial x(T)}\right)^T$ is a $2n \times r$ matrix of the same rank as $\frac{\partial O}{\partial x(T)}$ and $O(x(T)) - O_t$ an $r \times 1$ vector. By the rank-nullity theorem, $\text{rank}\left(\frac{\partial O}{\partial x(T)}\right)^T + \text{nullity}\left(\frac{\partial O}{\partial x(T)}\right)^T = r$, so the nullity of $\left(\frac{\partial O}{\partial x(T)}\right)^T = r - k \geq 1$. Thus, there exists nonzero $O(x(T)) - O_t$ such that $\left(O(x(T)) - O_t\right)^T \frac{\partial O}{\partial x(T)} = 0$. It is possible, however, for the required $O(x(T))$ to be unreachable by any $x(T)$ in the phase space or for $x(T)$ to represent a local maximum or saddle point.

Define the reachable set R to be the set of $x(T)$ such that there exists a control history $\epsilon(t)$ producing a final system state $x(T)$, with fixed initial conditions $x(0)$ and final time T. Then in order for all locally optimal ϵ to be globally optimal, for any $\zeta(T) \in R$ such that $O(\zeta(T)) \neq O_t$, $O(\zeta(T)) - O_t \notin N\left(\left(\frac{\partial O}{\partial x(T)}\right)^T |_{\zeta(T)}\right)$, where $N(\cdot)$ denotes the nullspace of a matrix. The geometric intuition is developed below.

Suppose that $x(T)$ were such that $\left(O(x(T)) - O_t\right)$ is in the nullspace of the transposed Jacobian $\left(\frac{\partial O}{\partial x(T)}\right)^T$, i.e. $\left(O(x(T)) - O_t\right)$ is in the cokernel of the Jacobian. This is just the quotient space $O(R)/Im(A)$, where A is the Jacobian at x. Note that since $\left(O(x(T)) - O_t\right)$ is in the cokernel of the Jacobian, it is not in the image of the Jacobian. Defining the differntiable manifold M to be $O(R)$, $\left(O(x(T)) - O_t\right)$ is orthogonal to the tangent space of M at the point $O(x(T))$. Note that $M = O(R)$ is a differentiable manifold if R is an open set, O is differentiable (which I have implicitly assumed) and O is a local homeomorphism. The assumption that $\frac{\delta x(T)}{\delta \epsilon}$ is surjective implies that R is open. Thus, a sufficient condition for no local minima is that $O(\zeta(T)) - O_t$ not be orthogonal to the tangent space of $O(R)$ at $\zeta(T)$. For instance, if $O(R)$ were a sphere centered at O_t, this condition would not be satisfied (the radii of a circle are perpendicular to the tangent space).

The orthogonality condition on O is somewhat stringent and non-intuitive, so let us check the Hessian matrix to see if all local minima are global minima, i.e. that critical points not at global minima correspond to local maxima or saddle points. Critical points that are not local minima offer no trap for the gradient algorithm and may be ignored. To that end, differentiating (12) yields

$$\frac{\delta J}{\delta \epsilon(t) \delta \epsilon(t')} = \frac{\partial J}{\partial x(T)} \frac{\delta x(T)}{\delta \epsilon(t) \delta \epsilon(t')} + \frac{\delta}{\delta \epsilon(t')} \left(\frac{\partial J}{\partial x(T)}\right) \frac{\delta x(T)}{\delta \epsilon(t)}. \qquad (17)$$

The second term in the summand may be expanded to yield

$$\frac{\delta J}{\delta \epsilon(t) \delta \epsilon(t')} = \frac{\partial J}{\partial x(T)} \frac{\delta x(T)}{\delta \epsilon(t) \delta \epsilon(t')} + \left[\frac{\partial^2 J}{\partial x(T)^2} \frac{\delta x(T)}{\delta \epsilon(t')}\right]^T \frac{\delta x(T)}{\delta \epsilon(t)}. \qquad (18)$$

Since $\frac{\partial J}{\partial x(T)} = 0$ from above, the first term in the summand goes to 0 and

$$\frac{\delta J}{\delta\epsilon(t)\delta\epsilon(t')} = \left[\frac{\partial^2 J}{\partial x(T)^2}\frac{\delta x(T)}{\delta\epsilon(t')}\right]^T \frac{\delta x(T)}{\delta\epsilon(t)} = \left(\frac{\delta x(T)}{\delta\epsilon(t')}\right)^T \frac{\partial^2 J}{\partial x(T)^2}\frac{\delta x(T)}{\delta\epsilon(t)}, \qquad (19)$$

using the fact that $\frac{\partial^2 J}{\partial x(T)^2}$ is a symmetric matrix (this follows from twice continuous differentiability of J with respect to $x(T)$). This is same expression as for the quantum mechanical Hessian in [9]. Local minima then correspond to those $\epsilon(t)$ for which $\left(\frac{\delta x(T)}{\delta\epsilon(t')}\right)^T \frac{\partial^2 J}{\partial x(T)^2}\frac{\delta x(T)}{\delta\epsilon(t)}$ has eigenvalues ≥ 0. In fact, one can show that only a finite number of these eigenvalues are nonzero. This conclusion follows using the same argument as for quantum mechanical systems in [5].

4 Singular Controls

To analyze singular controls, consider the Pontryagin maximum (or minimum) principle. Let the dynamics of (6) above be given by $\dot{x} = F_0(x(t)) + u(t)F_1(x(t))$, where x is the state and $u = \epsilon$ the control. Define the Pontryagin Hamiltonian

$$\overline{H}(t) = \pi^T(t)\big(F_0(x(t)) + u(t)F_1(x(t))\big), \qquad (20)$$

where π is analogous to the Lagrange multiplier and satisfies the Euler-Lagrange equations (7-9). Then the Pontryagin maximum principle states that the cost functional J is minimized by a control ϵ_0 producing state $\overline{x}$ and $\overline{\pi}$ only if the Pontryagin Hamiltonian $\overline{H}$ satisfies

$$\overline{\pi}^T(t)\big(F_0(\overline{x}(t)) + \epsilon_0(t)F_1(\overline{x}(t))\big) \leq \overline{\pi}^T(t)\big(F_0(\overline{x}(t)) + \epsilon(t)F_1(\overline{x}(t))\big) \qquad (21)$$

for all times t, i.e. ϵ_0 minimizes $\overline{H}$. For simplicity, introduce the notation $z = [x, \pi]^T$ and $\overline{H}(t) = H_0(z(t)) + H_1(z(t))u(t)$. For singular controls, $H_1 = 0$, i.e. $\frac{\delta\overline{H}}{\delta u}$ does not depend on u. (In fact, this occurs even for regular controls since all system dynamics considered here are linear in u, from (2-4)). In order to solve for singular controls, then, one differentiates the equation $H_1(z(t)) = 0$ with respect to time [1], [9]. The differentiation then yields another linear equation in ϵ, and can be solved for ϵ if the ϵ coefficient is nonzero on an open dense subset of the time interval $[0, T]$. If that occurs, one can differentiate the coefficient and constant term with respect to time again, setting each derivative equal to 0. Roughly speaking, the number of times it is necessary to differentiate equals the order of the resulting control. More precisely, ϵ is of order q if

$$\exists\, \overline{\alpha} = (1, i_1, \ldots, i_{q-1}, 1) \in \{0,1\}^{q+1} \text{ s.t. } A(\overline{\alpha}) \text{ is dense in } [T_1, T_2], \qquad (22)$$

where $A(\overline{\alpha}) = \{t \in [T_1, T_2] \mid H_{\overline{\alpha}}(z(t)) \neq 0\}$ and $\mid \alpha \mid$ is the number of elements in the vector α. Here $H_{1i_1i_2\ldots i_q}$ is defined recursively by $H_{1i_1i_2\ldots i_q} = \big\{H_{1i_1i_2\ldots i_{q-1}}, H_{i_q}\big\}$, where $\{\cdot,\cdot\}$ is the Poisson bracket. This algorithm is analogous to the one for tracking control proposed in [2], where the control field is designed so that a system observable follows a given path.

5 Conclusion and Future Work

Numerical simulations show that classical mechanical systems appear to behave like quantum mechanical ones: in evolving an optimal control field for a target state, a gradient algorithm never gets stuck. However, this result has not been proved generally for classical systems and only applies to systems in pure states. In particular, [4] uses a cost functional $J = \int_\Omega f(\omega)\rho(\omega)\, d\omega$, where $\rho(\epsilon)$ is a probability distribution of states, to show that a local minimum of J with respect to ρ is always a global minimum. One interesting area of future work would be linking this conclusion to the Hamiltonian dynamics of a classical system and constructing the analogy with a mixed-state quantum system, in which traps may exist [3].

Another interesting area of future work is that of singular control. Even in the quantum mechanical case, little is known about singular controls, and their optimality for both types of systems should be interesting to study. Numerical quantum mechanical experiments have shown no traps, but this result has not be proven for general systems. It would also be interesting to investigate the relative densities of singular and regular controls; numerically, regular controls appear to be quite closely spread throughout the control landscape, but singular controls are very hard to find without explicitly searching for them.

References

[1] B. Bonnard and M. Chyba. *Singular trajectories and their role in control theory*. Springer Verlag, 2003.

[2] Y. Chen, P. Gross, V. Ramakrishna, H. Rabitz, K. Mease, and H. Singh. Control of classical regime molecular objectives–Applications of tracking and variations on the theme* 1. *Automatica*, 33(9):1617–1633, 1997.

[3] M. Hsieh, R. Wu, and H. Rabitz. Topology of the quantum control landscape for observables. *The Journal of chemical physics*, 130:104109, 2009.

[4] Alexander Pechen and Herschel Rabitz. Unified analysis of control in classical and quantum systems. 2010.

[5] H. Rabitz, T.S. Ho, M. Hsieh, R. Kosut, and M. Demiralp. Topology of optimally controlled quantum mechanical transition probability landscapes. *Physical Review A*, 74(1):12721, 2006.

[6] H.A. Rabitz, M.M. Hsieh, and C.M. Rosenthal. Quantum optimally controlled transition landscapes. *Science*, 303(5666):1998, 2004.

[7] C.D. Schwieters and H. Rabitz. Optimal control of nonlinear classical systems with application to unimolecular dissociation reactions and chaotic potentials. *Physical Review A*, 44(8):5224–5238, 1991.

[8] C.D. Schwieters and H. Rabitz. Optimal control of classical systems with explicit quantum-classical-difference reduction. *Physical Review A*, 48(4):2549–2557, 1993.

[9] Rebing Wu, Jason Dominy, Tak-San Ho, and Herschel Rabitz. The Role of Singular Controls in Optimizing Quantum Dynamics. 2010.