Duality and dynamics in Hamilton-Jacobi theory for fully convex problems of control

R.Tyrrell. Rockafellar and Peter R. Wolenski

Abstract

This paper describes some recent results in Hamilton-Jacobi theory that hold under strong convexity assumptions on the data. Generalizations of linear-quadratic control models satisfy such assumptions, for example. The results include a global method of characteristics and a strong duality theory.

1 Introduction

Convexity assumptions in optimization and optimal control theory play a role analogous to the role of linearity in functional analysis and ordinary differential equation theory. This is underlined by two themes: (1) properties that hold locally in general cases have global versions under convexity assumptions, and (2) the presence of convexity offers dual problem formulations that unite important concepts and offer symmetrical results. This paper and its companion summarize recent results in [1] and [2] on how strong convexity assumptions can be exploited in Hamilton-Jacobi theory.

1.1 Problem formulation

Many important issues in optimal control and the calculus of variations revolve around a value function $V : [0,\infty) \times I\!R^n \to \overline{I\!R} := [-\infty, \infty]$ of the type

$$\begin{aligned} V(\tau,\xi) \quad &:= \inf\Big\{ g\big(x(0)\big) + \textstyle\int_0^\tau L\big(x(t),\dot{x}(t)\big)dt \,\Big|\, x(\tau)=\xi \Big\}, \\ V(0,\xi) \quad &= g(\xi), \end{aligned}$$

which propagates an initial cost function $g : I\!R^n \to \overline{I\!R}$ forward from time 0 in a manner dictated by a Lagrangian function $L : I\!R^n \times I\!R^n \to \overline{I\!R}$. The possible extended-real-valuedness of g and L serves in the modeling of the constraints and dynamics involved in this propagation, as will be illustrated by an example in Section 2. The minimization takes place over the arc space $\mathcal{A}^1_n[0,\tau]$ of absolutely continuous functions that map into $I\!R^n$.

The complete technical assumptions will be given in Section 4, but we emphasize that convexity will play a major role in our results. In particular we shall assume L is convex *jointly* in the variables (x, v), and g will also be assumed to be convex. We call such variational problems *Fully Convex* (FC).

The theoretical advantage of the FC formulation comes from the fact that it closely resembles a calculus of variations problem, and thus results from that extensive and classical subject provide a roadmap for investigation. Assumptions in the state variable can be easily formulated and brought to the forefront, since the FC formulation focuses on finding only the optimal arcs. If the problem comes from an optimal control model, the issue of capturing other control information such as a feedback law is isolated to another stage of the process. Once the optimal velocities are known, however, finding the optimal control reduces to a finite-dimensional optimization problem.

1.2 The Hamilton-Jacobi equation

If V is smooth, then it satisfies the Hamilton-Jacobi (HJ) equation

$$\tfrac{\partial}{\partial\tau} V(\tau,\xi) = -H\big(\xi, \nabla_\xi V(\tau,\xi)\big), \qquad (1)$$

where $H : I\!R^n \times I\!R^n \to I\!R$ is the Hamiltonian defined by

$$H(x,y) = \sup_{v\in I\!R^n} \Big\{ \langle y, v\rangle - L(x,v) \Big\}.$$

It is well appreciated that smoothness of solutions to (1) will only happen under very special circumstances, but that a general solution concept can nonetheless be used to characterize V as the solution to (1). It is also well appreciated that numerical methods that approximate the nonsmooth solution are difficult to implement practically. The Hamilton-Jacobi equation arising in the FC model is a nonlinear first-order PDE that has particular structure that has not yet been fully exploited.

2 Notable example: linear-quadratic control with hard control constraints

Perhaps the most widely-used state-based optimal control model is the Linear-Quadratic Regulator (LQR). This problem has quadratic costs and linear dynamics, and is of the form

$$\inf\Big\{ x(0)^\top P x(0) + \int_0^\tau \big\{ x(t)^\top Q x(t) + u(t)^\top R u(t) \big\}\, dt \Big\}$$

where the minimization is over $x(\cdot)$ in $\mathcal{A}_n^1[0,\tau]$ with $x(\tau)=\xi$ and measurable $u(\cdot):[0,\tau]\to I\!R^m$ that satisfy

$$\dot{x}(t) = Ax(t) + Bu(t) \quad \text{a.e. } t\in[0,\tau]. \qquad (2)$$

The matrices A, B, P, Q, and R are all of the appropriate dimension with P, Q, and R positive definite.

The LQR problem can be equivalently described in our problem formulation by setting

$$L(x,v) := \inf\Big\{x^\top Qx + u^\top Ru : \; v = Ax+Bu\Big\}. \quad (3)$$

Then for a given $x(\cdot)\in\mathcal{A}_n^1[0,\tau]$, one has

$$\int_0^\tau L\big(x(t),\dot{x}(t)\big)\,dt < \infty$$

if and only if there exists a measurable $u(\cdot)$ satisfying (2), and moreover, the infimum and the optimal state trajectories in the FC problem will coincide with those in the LQR problem. Note that L as given in (3) is convex in (x,v), although it may not be differentiable nor finite everywhere (by convention, the inf over the empty set is $+\infty$). Finally, if an optimal velocity $\bar{v}$ is found at a particular time and state $\bar{x}$, then an optimal control $\bar{u}$ is a point achieving the inf in (3) that defines $L(\bar{x},\bar{v})$.

The LQR formulation's main advantage is the ease and manner in which solutions can be obtained. In fact, there exists a differentiable map $K:[0,\tau]\to\mathcal{M}_{n\times m}$ so that a feasible pair $\big(x(\cdot),u(\cdot)\big)$ solves the LQR problem if and only if $x(t)=K(t)u(t)$. The matrix-valued map K is found through solving a Riccati equation, which can be done relatively easily, at least in low dimensions. Note that this *characterizes* the optimal solutions.

The FC problem formulation captures important features of LQR while introducing much greater flexibility. We illustrate by considering a natural generalization of LQR by simply adding a control constraint $u(t)\in U$ to (2), where $U\subseteq I\!R^m$ can be taken closed and convex. Such *hard* constaints on the control function $u(\cdot)$ are typical in applications where bounds on the input are unavoidable. The methodology of LQR now breaks down, and in particular there is no longer an obvious utilization of the solution to a Riccati equation, nor is a *linear* feedback law to be expected. But note if the constraint $u\in U$ is added to the infimum in (3), the resulting function L is still jointly convex, and hence can be treated in the FC framework.

If recourse to solutions of the Riccati equation is no longer available, then the Hamilton-Jacobi equation must be dealt with directly, an intimidating task for general nonlinear problems. But, as we shall see, the FC case has very special features that potentially open up a much greater utilization of the HJ equation.

3 Convex Analysis and Duality

After a brief review of concepts from convex analysis, this section introduces the dual data and its problem formulation, and then states results obtained by Rockafellar from the 1970's. The latter results hold under the assumptions that will be imposed in Section 4.

Suppose $f: I\!R^n\to\overline{I\!R}$ is convex, lower semicontinuous, and proper. The subgradient set $\partial f(\bar{x})$ of f at a point $\bar{x}$ is the set of vectors $y\in I\!R^n$ that satisfy

$$f(x)\geq f(\bar{x}) + \langle y, x-\bar{x}\rangle \qquad \forall x\in I\!R^n.$$

Associated with f is its Legendre-Fenchel conjugate $f^*: I\!R^n\to\overline{I\!R}$, another convex, lower semicontinuous, and proper function, given by

$$f^*(y) = \sup_{x\in I\!R^n}\big\{\langle y,x\rangle - f(x)\big\}.$$

The Legendre-Fenchel transform f^{**} of f^* merely recovers f, and thus convex functions naturally come in pairs. Furthermore the subgradients of f and f^* are related by

$$y\in\partial f(x) \quad\Leftrightarrow\quad x\in\partial f^*(y),$$

which says that the subgradient mappings of conjugates are inverse to each other.

Notice that the Hamiltonian H as a function of its second argument is the Legendre-Fenchel transform of the second argument of L. In FC theory, where convexity of L in both variables is assumed, the full conjugate of L also plays a significant role. The *dual Lagrangian* $\tilde{L}(y,w)$ is defined as $L^*(w,y)$:

$$\tilde{L}(y,w) := \sup_{(x,v)\in I\!R^{2n}}\Big\{\langle w,x\rangle + \langle y,v\rangle - L(x,v)\Big\}$$

A certain symmetry in the assumptions on L and $\tilde{L}$ will be pointed out in the next section. Of course $\tilde{L}$ has its own Hamiltonian $\tilde{H}$, but it turns out that

$$\tilde{H}(y,x) = -H(x,y), \qquad (4)$$

and so this is not a new piece of data entering the picture. Joint convexity in L corresponds to $x\mapsto H(x,y)$ being concave for each fixed y, and $y\mapsto H(x,y)$ convex for each fixed x, and our assumptions will imply H is always finite. For such H, $\partial H(x,y)$ denotes the Clarke generalized gradient in both variables (x,y), but this simplifies in the special case of concave/convex H to the product

$$\partial H(x,y) = \partial_x H(x,y)\times\partial_y H(x,y),$$

where $\partial_y H(x,y)$ is the convex subgradient in y and $\partial_x H(x,y)$ is the concave supergradient in x (= the negative of the convex subgradient of the map $x \mapsto -H(x,y)$).

Just as L and g gave rise to a so-called *primal* variation problem that is used to define V, the data $\tilde{L}$ and $\tilde{g}(\eta) := g^*(\eta)$ give rise to a *dual* variational problem, defined by

$$\begin{aligned}\tilde{V}(\tau,\eta) &:= \inf\Big\{\tilde{g}\big(y(0)\big) + \textstyle\int_0^\tau \tilde{L}\big(y(t),\dot{y}(t)\big)dt \,\Big|\, y(\tau)=\eta\Big\},\\ \tilde{V}(0,\eta) &= \tilde{g}(\eta),\end{aligned}$$

This problem is not the dual problem as introduced in [4], but is closely related and uses the same data.

Rockafellar in a series of papers [3]-[7] developed a theory under full convexity assumptions. These results mainly focused on optimality conditions, an existence theory, and duality, and play a major role in understanding the nature of the FC variation problem and in proving the main results stated in this paper. They are summarized in the following two theorems.

Theorem 3.1 (Rockafellar)
The following are equivalent for a given pair of arcs $x(\cdot)$ and $y(\cdot)$ satisfying $x(\tau)=\xi$ and $y(\tau)=\eta$.

(a) $x(\cdot)$ solves the primal problem and $y(\cdot)$ solves the dual problem.

In (b)-(e), $(x(\cdot),y(\cdot))$ in addition are assumed to satisfy the transversality condition $y(0) \in \partial g(x(0))$, which has the equivalent form $x(0) \in \partial\tilde{g}(y(0))$.

(b) $y(\cdot)$ is a multiplier for $x(\cdot)$ that solves the primal problem:

$$\big(\dot{y}(t),y(t)\big) \in \partial_{x,v}L\big(x(t),\dot{x}(t)\big).$$

(c) $x(\cdot)$ is a multiplier for $y(\cdot)$ that solves the dual problem:

$$\big(\dot{x}(t),x(t)\big) \in \partial_{y,w}\tilde{L}\big(y(t),\dot{y}(t)\big).$$

(d) $(x(\cdot),y(\cdot))$ is a primal Hamiltonian trajectory:

$$\big(-\dot{y}(t),\dot{x}(t)\big) \in \partial_{x,y}H\big(x(t),y(t)\big).$$

(e) $(y(\cdot),x(\cdot))$ is a dual Hamiltonian trajectory:

$$\big(-\dot{x}(t),\dot{y}(t)\big) \in \partial_{y,x}\tilde{H}\big(y(t),x(t)\big).$$

Another key result under our convexity assumptions is the constancy of the Hamiltonian:

Theorem 3.2 ([3]) *For $\big(x(\cdot),y(\cdot)\big)$ as in the last Theorem, the map $t \mapsto H\big(x(t),y(t)\big)$ is constant.*

4 Assumptions

We now state precisely the assumptions that are in effect throughout the paper. We first have the

Endpoint Assumption.

(A0) *The initial function g is convex, proper and lower semicontinuous on $\mathbb{R}^n$.*

If one of the functions L, $\tilde{L}$, H, and $\tilde{H}$ are known, then all the other functions are completely determined. Thus the assumptions encoding dynamic constraints and running costs can be given for any one of the four functions, and it is worth seeing the various relationships among them since one form may be more easily verfiable in a particular case.

4.1 Assumptions - Lagrangian form

We shall assume

(A1) The Lagrangian function L is convex, proper and lower semicontinuous on $\mathbb{R}^n \times \mathbb{R}^n$.

(A2) The set $F(x) := \operatorname{dom} L(x,\cdot)$ is nonempty for all x, and there is a constant ρ such that $\operatorname{dist}(0,F(x)) \le \rho\big(1+|x|\big)$ for all x.

(A3) There are constants α and β and a coercive, proper, nondecreasing function θ on $[0,\infty)$ such that $L(x,v) \ge \theta\big(\max\{0, |v|-\alpha|x|\}\big) - \beta|x|$ for all x and v.

The dual Lagrangian also satisfies a set of assumptions ($\tilde{\text{A}}$1)-($\tilde{\text{A}}$3) equivalent to (A1)-(A3), and these are being labelled in the same manner as above except with L replaced by $\tilde{L}$. General considerations immediately give that ($\tilde{\text{A}}$1) is equivalent to (A1), but our assumptions are tailored in such a manner that further symmetry is maintained. In fact, L satisfies (A2) (resp. (A3)) if and only if $\tilde{L}$ satisfies ($\tilde{\text{A}}$3) (resp. ($\tilde{\text{A}}$2)).

4.2 Assumptions - Hamiltonian form

It is sometimes easier to verify the assumptions in terms of the Hamiltonian, and so we state these next. There is a direct ordered relationship with the Lagrangian form given above, with (A1) and (a1) equivalent, and similarly with (A2) and (a2), (A3) and (a3).

(a1) The Hamiltonian function $H(x,y)$ is everywhere finite on $\mathbb{R}^n \times \mathbb{R}^n$, concave in x, convex in y.

(a2) There are constants α and β and a finite, convex function φ such that

$$H(x,y) \le \varphi(y) + (\alpha|y|+\beta)|x| \quad \text{for all } x,\ y.$$

(a3) There are constants γ and δ and a finite, concave function ψ such that

$$H(x,y) \ \geq \ \psi(x) - (\gamma|x| + \delta)|y| \quad \text{for all } x, \ y.$$

In view of (4), analogous dual Hamiltonian assumptions on $\tilde{H}$ have the same relationship to (a1)-(a3) as the dual Lagrangian assumptions ($\tilde{\text{A}}$1)-($\tilde{\text{A}}$3) have to (A1)-(A3).

4.3 Example

Suppose L is given as in (3), then H has the form

$$\begin{aligned} H(x,y) &= \langle y, Ax\rangle - x^\top Qx + \sup_u \Big\{\langle y, Bu\rangle - u^\top Ru\Big\} \\ &= \langle y, Ax\rangle - x^\top Qx + \tfrac{1}{4} y^\top BR^{-1}B^\top y. \end{aligned}$$

It is an easy matter to verify that (a1)-(a3) hold in this case. If the hard constraint $u \in U$ is added to the data, then this constraint is added to the sup in the above formula, and H has the form

$$\begin{aligned} H(x,y) = \ & \langle y, Ax\rangle - x^\top Qx + \tfrac{1}{4} y^\top BR^{-1}B^\top y \\ & -\tfrac{1}{4} \inf_{u\in U} \|u - R^{-1}B^\top y\|_R^2, \end{aligned}$$

where $\|w\|_R^2 := w^\top Rw$. It is immediate that H in this case still satisfies (a1)-(a3).

5 Main results

We now turn to the new results about the value function V that are consequences of all the hypothesized convexity. The first result contains some structural regularity properties.

Theorem 5.1 *For each fixed $\tau \geq 0$, the value function $V_\tau(\xi) := V(\tau, \xi)$ is convex, proper, and lower semicontinuous. Moreover, V_τ depends epi-continuously on τ.*

The epi-continuity statement means that whenever $\tau_i \to \tau$ as $i \to \infty$ with $\tau_i \geq 0$, one has

$$\liminf_{i\to\infty} V(\tau_i, \xi_i) \geq V(\tau, \xi)$$

for every sequence $\xi_i \to \xi$, and

$$\limsup_{i\to\infty} V(\tau_i, \xi_i) \leq V(\tau, \xi)$$

for some sequence $\xi_i \to \xi$.

The next two results are the striking features of FC problems that have no analogue in general nonlinear theory. They assert that the subgradients and function values of the value function V can be found by a global "method of characteristics". We emphasize that this method operates globally for $\tau \geq 0$. For general nonlinear problems, the characteristic method will be valid only locally, and it is usually difficult to estimate the size of the region where it is operational.

We have already encountered Hamiltonian trajectories in the optimality conditions in Theorem 3.1. In a direct analogy to classical Hamilton-Jacobi theory, these same arcs are the characteristic curves that determine V and its subgradients. A little more notation helps to state the results. Define the reachable set $S_\tau(\xi, \eta)$ at time $\tau \geq 0$ emanating from an initial point $(\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^n$ as the set of terminal points $(\xi', \eta') \in \mathbb{R}^n \times \mathbb{R}^n$ of a Hamiltonian trajectory. To be precise, $(\xi', \eta') \in S_\tau(\xi, \eta)$ if and only if there exists an arc $\big(x(\cdot), y(\cdot)\big)$ satisfying

$$\begin{aligned} -\dot{y}(t) &\in \partial_x H\big(x(t), y(t)\big), \\ \dot{x}(t) &\in \partial_y H\big(x(t), y(t)\big), \end{aligned}$$

and

$$\big(x(0), y(0)\big) = (\xi, \eta) \quad \text{and} \quad \big(x(\tau), y(\tau)\big) = (\xi', \eta').$$

The following result says that the graph of the subgradient mapping of g is propagated into the subgradient mapping of V_τ by Hamiltonian trajectories.

Theorem 5.2 *For all $\tau \geq 0$, one has*

$$S_\tau\big(\operatorname{gph} \partial g\big) = \operatorname{gph} \partial V_\tau.$$

That is, for a Hamiltonian trajectory $\big(x(\cdot), y(\cdot)\big)$ defined on $[0, \tau]$, one has $y(0) \in \partial g(x(0))$ if and only if $y(\tau) \in \partial V_\tau(x(\tau))$.

The next result requires further concepts of subgradient beyond that of convex analysis in order to "differentiate" V in both variables (τ, ξ) jointly. Certainly V will hardly ever be convex in (τ, ξ), and for this reason HJ theory was not developed for FC earlier in the 70's - further developments in variational analysis were required.

Consider any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and let x be any point at which $f(x)$ is finite. A vector $y \in \mathbb{R}^n$ is a *regular subgradient* of f at x, written $y \in \hat{\partial} f(x)$, if

$$f(x') \ \geq \ f(x) + \langle y, \, x' - x\rangle + o(|x' - x|).$$

It is a (*general*) *subgradient* of f at x, written $y \in \partial f(x)$, if there is a sequence of points $x_i \to x$ with $f(x_i) \to f(x)$ for which regular subgradients $y_i \in \hat{\partial} f(x_i)$ exist with $y_i \to y$. For a convex function f, $\hat{\partial} f(x)$ and $\partial f(x)$ reduce to the convex subgradient, and hence there is no ambiguity in using the same notation. In the case of the value function V, the "partial subgradient" notation

$$\partial_\xi V(\tau, \xi) = \big\{\eta \,\big|\, \eta \in \partial V_\tau(\xi)\big\}$$

can be interpreted equally in any of the senses above.

Theorem 5.3 *The subgradients of V on $(0,\tau)\times I\!R^n$ have the property that*

$$(\sigma,\eta)\in\partial V(\tau,\xi) \iff (\sigma,\eta)\in\hat{\partial}V(\tau,\xi) \iff \eta\in\partial_\xi V(\tau,\xi),\quad \sigma=-H(\xi,\eta).$$

In particular, therefore, V satisfies the generalized Hamilton-Jacobi equation:

$$\sigma=-H(\xi,\eta) \quad \text{for all} \quad (\sigma,\eta)\in\partial V(\tau,\xi)$$

when $\tau>0$.

Again one may note the equivalence of the above properties, which is another hallmark of convexity.

The natural question of whether V is uniquely determined as the solution to the generalzed HJ equation (plus boundary conditions) was only resolved recently by Galbraith [8], [9]. One of the achievements of viscosity theory was the resolution to this uniqueness issue under certain situations, but our assumptions are not covered by that theory. Different techniques were developed in [8], [9] to prove uniqueness in which full convexity assumptions are exploited.

6 Value function conjugacy

This last section contains results that further illustrate the rich structure of the FC formulation. First note that L and $\tilde{L}$ satisfy the same assumptions, and thus the results stated above for V are equally valid for $\tilde{V}$. Recall that $V_\tau(\cdot):=V(\tau,\cdot)$ and $\tilde{V}_\tau(\cdot):=\tilde{V}(\tau,\cdot)$ are convex, and the next result states that in fact they are conjugate.

Theorem 6.1 *For each $\tau\geq 0$, one has*

$$V_\tau^*(\eta)=\tilde{V}_\tau(\eta) \quad \text{for all} \quad \eta\in I\!R^n.$$

Many attributes that are imposed on convex functions have counterpart dual properties, and thus the previous theorem offers a convenient tool to derive properties of V through those of $\tilde{V}$. As an illustration, differentiability of a finite convex function f is equivalent to strict convexity of f^* on the set where $\{y:\partial f^*(y)\neq\emptyset\}$. Now if $\tilde{L}$ is strictly convex, it is easy to show $\tilde{V}_\tau$ is strictly convex. This simple observation and Theorem 6.1 leads immediately to the following regularity result.

Corollary 6.2 *If L is finite and differentiable, then V is C^1 on $[0,\infty)\times I\!R^n$ (for any g).*

Another illustration is the propagation of finiteness and coercivity. Recall that a function $\theta: I\!R^m\to\overline{I\!R}$ is coercive by definition if $\lim_{|z|\to\infty}\theta(z)/|z|=+\infty$. The coercivity of a convex function f is equivalent to f^* being finite everywhere.

Our assumptions are such that the following proposition is relatively easy to prove.

Proposition 6.3 *(a) If g is finite on $I\!R^n$, then V is finite on $[0,\infty)\times I\!R^n$.*

(b) If L is finite on $I\!R^n\times I\!R^n$, then V is finite on $(0,\infty)\times I\!R^n$.

The next corollary now follows immediately from Theorem 6.1 and Proposition 6.3.

Corollary 6.4 *(a) If g is coercive on $I\!R^n$, then V_τ is coercive on $I\!R^n$ for all $\tau\geq 0$.*

(b) If L is coercive on $I\!R^n\times I\!R^n$, then V_τ is coercive on $I\!R^n$ for all $\tau\geq 0$.

References

[1] R.T. Rockafellar and P.R. Wolenski, "Convexity in Hamilton-Jacobi theory I: dynamics and duality," SIAM J. Control and Opt. 39 (2001), 1323–1350.

[2] R.T. Rockafellar and P.R. Wolenski, "Convexity in Hamilton-Jacobi theory II: envelope representations," SIAM J. Control Opt. 39 (2001), 1351-1372.

[3] R. T. Rockafellar, "Generalized Hamiltonian equations for convex problems of Lagrange," Pacific J. Math. 33 (1970), 411–428.

[4] R. T. Rockafellar, "Existence and duality theorems for convex problems of Bolza," Trans. Amer. Math. Soc. 159 (1971), 1–40.

[5] R. T. Rockafellar, "Semigroups of convex bifunctions generated by Lagrange problems in the calculus of variations," Math. Scandinavica 36 (1975), 137–158.

[6] R.T. Rockafellar, "Linear-quadratic programming and optimal control," SIAM J. Control Opt. 25 (1987), 781–814.

[7] R.T. Rockafellar, "Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs," SIAM J. Control Opt. 27 (1989), 1007–1025.

[8] G. Galbraith, "Applications of variational analysis to optimal trajectories and nonsmooth HJ theory," Ph.D. Thesis, University of Washington, 1999.

[9] G. Galbraith, "The role of cosmically Lipschitz mappings in nonsmooth HJ theory," preprint.