Optimal Control Theory

Mrinal Kanti Ghosh
Department of Mathematics
Indian Institute of Science
Bangalore 560 012
email : mkg@math.iisc.ernet.in

1 Introduction

Optimal Control Theory deals with optimization problems involving a controlled dynamical system. A controlled dynamical system is a dynamical system in which the trajectory can be altered continuously in time by choosing a control parameter $u(t)$ continuously in time. A (deterministic) controlled dynamical system is usually governed by an ordinary differential equation of the form

$$\left.\begin{array}{rcl} \dot{x}(t) & = & f(t, x(t),\ u(t)), t > 0 \\ \\ x(0) & = & x_0 \in \mathbb{R}^d \end{array}\right\} \tag{1.1}$$

where $u : \mathbb{R}^d \to U$ is a function called the control, U a given set called the control set, and

$$f : \mathbb{R}_+ \times \mathbb{R}^d \times U \to \mathbb{R}^d.$$

By choosing the value of $u(t)$, the state trajectory $x(t)$ can be controlled. The objective of controlling the state trajectory is to minimize a certain cost associated with (1.1).

Consider the following important problem studied by Bushaw and Lasalle in the field of aerospace engineering. Let $x(t)$ represent the deviation of a rocket trajectory from some desired flight path. Then, provided the deviation is small, it is governed to a good approximation by

$$\left.\begin{array}{rcl} \dot{x}(t) & = & A(t)x(t) + B(t)u(t), t > 0 \\ x(0) & = & x_0, \qquad\qquad\qquad x(T) = 0, \\ & & u(t) \in [-1, 1], \end{array}\right\} \tag{1.2}$$

where A and B are appropriate matrix valued functions. The vector x_0 is the initial deviation, $u(t)$ is the rocket thrust at time t, and T is the final time. The problem is that of finding a rocket thrust history to minimize

$$J(u) := \int_0^T c(t, x(t), u(t))dt \tag{1.3}$$

where c is a certain cost function. Here the aim is to reduce to deviation to zero over the time interval [0,T], and at the same time, to make the value of $J(u)$ as small as possible. The cost function c is chosen so that $J(u)$ has the interpretation of, say, fuel consumption or (via a change of independent variable) the time taken to reduce the deviation to zero.

Subsequently Pontryagin and his collaborator studied the following optimal control problem : minimize

$$\left.\begin{array}{c} \displaystyle\int_0^T c(t,x(t),u(t))dt \ + \ g(x(T)) \\ \\ \text{Subject} \quad \text{to} \\ \\ \dot{x}(t) = f(t,x(t),u(t)) \\ \\ x(0) \in C_0, \quad x(T) \in C_1, u(t) \in U \end{array}\right\} \tag{1.4}$$

where c, f, g are appropriate functions, and $C_0 \subset \mathbb{R}^d, C_1 \subset \mathbb{R}^d$. Minimization is carried out over control functions $u(\cdot)$ and the corresponding solutions to the differential equations which satisfy the constraints. The importance of the formulation of the optimal control problem cannot be overemphasized. It incorporates a large number of dynamic optimization problems of practical interest. In particular it also subsumes the classical problem of calculus of variations. Indeed, put

$$f(t,x,u) = u, \qquad g = 0.$$

Then the above problem reduces to: minimize

$$\int_0^T c(t,x(t),\dot{x}(t))dt$$

subject to

$$x(0) \in C_0, \quad x(T) \in C_1.$$

Thus we recover, as a special case of (1.4), the basic problem of calculus of variation which is of fundamental importance in Classical Mechanics and various other fields.

2 Optimal Control Problems

Let $U \subset \mathbb{R}^m$ be compact and

$$f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d.$$

Let $x(\cdot) : \mathbb{R}_+ \to \mathbb{R}^d$ be the solution of the controlled ordinary differential equation given by

$$\left.\begin{array}{rcl} \dot{x}(t) & = & f(t,x(t),u(t)), \ t > 0 \\ x(0) & = & x \in \mathbb{R}^d \end{array}\right\} \tag{2.1}$$

where $u(\cdot) : \mathbb{R}_+ \to \mathbb{R}^m$ is a measurable function such that $u(t) \in U$ *a.e.* Such a function is called a control function or simply control. We assume that

(A1) The function f is continuous, and Lipschitz in its second argument uniformly with respect to the first and third.

For a given control $u(\cdot)$, a Lipschitz continuous function $x(\cdot)$ satisfying (2.1) a.e. is called the state trajectory corresponding to $u(\cdot)$. Under the assumption $(A1)$, for a given control a unique state trajectory exists for all time for any initial condition $x(0) = x$.

Let

$$c : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$$

and

$$g : \mathbb{R}^d \to \mathbb{R}.$$

We assume that

(A2) The function c is continuous, and Lipschitz in its second argument uniformly with respect to the first and third.

(A3) The function g is Lipchitz continuous.

We now describe the typical optimal control problems.

(P1) Finite Horizon Problem : Let $T > 0$ be fixed.
Let

$$J_T(x, u(\cdot)) := \int_0^T c(t, x(t), u(t))dt + g(x(T))$$

i.e., $J_T(x, u(\cdot))$ is the total cost incurred over the finite time interval (horizon) $[0, T]$. Here the function c is referred to as the running cost and the function g is referred to as the terminal cost. Our objective is to minimize $J_T(x, u(\cdot))$ over all controls $u(\cdot)$.

A control $u^*(\cdot)$ is called an optimal control for the initial condition x for the finite horizon problem if

$$J_T(x, u^*(\cdot)) = \inf_{u(\cdot)} J_T\,(x, u(\cdot))$$

where the infimum above is over all controls . For given $\varepsilon > 0$, a control $u_\varepsilon(\cdot)$ is said to be an ε - optimal control if

$$J_T(x, u_\varepsilon(\cdot)) \leq \inf_{u(\cdot)} J_T(x, u(\cdot)) + \varepsilon.$$

Note that an optimal control is ε - optimal control for any $\varepsilon > 0$. On the other hand if $u^*(\cdot)$ is ε - optimal for every $\varepsilon > 0$, then it is optimal.

(P2) Exit Time Problem: Let $D \subset \mathbb{R}^d$ be a smooth, bounded domain. Let

$$\mathrm{T} = \inf\{t \geq 0| \;\; x(t) \notin \bar{D}\}.$$

Let

$$J_{\mathrm{T}}(x, u(\cdot)) := \int_0^{\mathrm{T}} c(t, x(t), u(t))dt + g(x(\mathrm{T})).$$

In the exit time problem we wish to minimize $J_{\mathrm{T}}(x, u(\cdot))$ over all controls.

(P3) Discounted Cost Problem: Let $\alpha > 0$ be the discount factor. The α- discounted cost on the infinite horizon is given by

$$J_\alpha(x, u(\cdot)) := \int_0^\infty e^{-\alpha t} c(t, x(t), u(t)) dt.$$

(P4) Ergodic Control Problem: In the ergodic control problem we seek to minimize the limiting time -average cost given by

$$\limsup_{T \to \infty} \frac{1}{T} \int_0^T c\,(t,\ x(t), u(t))\, dt.$$

3 Existence Theory

In this section we address the question of the existence of an optimal control. For this purpose we describe the finite horizon case in details. We first begin with the following example to demonstrate non-existence of optimal controls.

Example 3.1 Minimize

$$\int_0^1 x^2(t) dt$$

subject to

$$\dot{x} = u,\ x(0) = 0$$

$$u(\cdot) \in \{-1, 1\}.$$

Consider the sequence of controls $u_n(\cdot), n = 1, 2, \ldots,$ given by

$$u_n(t) = \begin{cases} 1 & \text{if } \ \frac{2k}{n} \leq t < \frac{2k+1}{n}, \\ \\ -1 & \text{if } \ \frac{2k+1}{n} \leq t < \frac{2k+2}{n}. \end{cases}$$

If $x_n(\cdot)$ denotes the corresponding state trajectories, then

$$\sup_n |\ x_n(t)| \leq n^{-1}.$$

Evidently

$$\int_0^1 x_n^2(t) dt \to 0 \qquad \text{as} \qquad n \to \infty.$$

So it follows that

$$\inf \int_0^1 x^2(t) dt = 0.$$

Now suppose there is an optimal control $u^*(\cdot)$, and $x^*(\cdot)$ the corresponding trajectory. Then

$$\int_0^1 x^{*2}(t)dt = 0$$

which implies $x^*(\cdot) \equiv 0$. Hence

$$\dot{x}^*(\cdot) \equiv 0.$$

but this is not tenable since $0 \notin \{-1, 1\}$.

We proceed to establish the existence of an optimal control. To this end we need an additional condition. Let

$$\tilde{f} : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^{d+1}$$

be defined by

$$\tilde{f}(t, x, u) = \begin{bmatrix} c(t, x, u) \\ f(t, x, u) \end{bmatrix}.$$

(A4) The set $\tilde{f}(t, x, U)$ is convex for all t, x.

Theorem 3.1 Assume $(A1) - (A4)$. Then there exists an optimal control.

Proof. We first observe that we can take $c \equiv 0$ without any loss of generality. Indeed, if c is present, we treat instead the problem

minimize $\quad \tilde{g}(\tilde{x}(T))$ subject to

$$\dot{\tilde{x}} = \tilde{f}(t, \tilde{x}, u)$$

where

$$\tilde{x} = [x_0, x]' \in \mathbb{R}^{d+1}, \qquad x_0(0) = 0$$

$$\tilde{g}(\tilde{x}) = x_0 + g(x).$$

Clearly the new problem is equivalent to the original one. This justifies the reduction. We claim that for any $x \in \mathbb{R}^d$, there exists a constant K such that

$$|\ x(t)\ | < K, \qquad \text{for all } t \in [0, T] \tag{3.1}$$

for any state trajectory $x(\cdot)$ under any control $u(\cdot)$, satisfying $x(0) = x$. Indeed, choose a fixed control $\bar{u}(\cdot)$, and let $\bar{x}(\cdot)$ be the corresponding trajectory with $\bar{x}(0) = x$. Set

$$z(t) = x(t) - \bar{x}(t).$$

Then for any $0 \le s \le t \le T$, we have

$$|\ z(t) - z(s)\ | \ \le \ \int_s^t |\ f(\sigma, x(\sigma), u(\sigma)) - f(\sigma, \bar{x}(\sigma), \bar{u}(\sigma))\ |\ d\sigma$$

$$
\begin{aligned}
&\leq \int_s^t \mid f(\sigma, x(\sigma), u(\sigma)) - f(\sigma, \bar{x}(\sigma), u(\sigma)) \mid d\sigma \\
&+ \int_s^t \mid f(\sigma, \bar{x}(\sigma), u(\sigma)) - f(\sigma, \bar{x}(\sigma), \bar{u}(\sigma)) \mid d\sigma \\
&\leq C_1 \int_s^t \mid z(\sigma) \mid d\sigma + C_2(t-s)
\end{aligned}
$$

where C_1 is the Lipschitz constant associated with f and C_2 is a uniform bound on the integrand of the second term. Note that C_2 depends on T, but it is a priori fixed. Thus there exists a constant C_3 such that

$$
\mid \mid z(t) \mid - \mid z(s) \mid \mid \quad \leq \quad C_3 \int_s^t (1+ \mid z(\sigma) \mid) d\sigma
$$

By Gronwall's inequality, $z(t)$ is uniformly bounded on $[0, T]$. Since $\bar{x}(\cdot)$ in a given process, it follows that then exists a $K > 0$ such that

$$
\mid x(t) \mid \quad \leq \quad K \qquad \forall t \in [0, T] \, .
$$

Now, let $\{u_n(\cdot)\}$ be a sequence of controls, and $\{x_n(\cdot)\}$ the corresponding trajectory with $x_n(0) = x$, such that

$$
g(x_n(T)) \to \inf g(x(T)) \tag{3.2}
$$

as $n \to \infty$, where the infimum on the right is over all controls. By $(A1)$ and Ascoli's theorem there exists a Lipschitz continuous function $x(\cdot)$ with $x(0) = x$, such that

$$
\sup_{t \in [0,T]} \mid x_n(t) - x(t) \mid \to 0 \tag{3.3}
$$

along a suitable subsequence. We next show that $x(\cdot)$ satisfies (2.1) for some control $u(\cdot)$. To achieve this we first show that for any $p \in \mathbb{R}^d$

$$
\sup_{u \in f(\sigma, x(\sigma), U)} \langle p, u \rangle \geq \langle p, \dot{x}(\sigma) \rangle \, , \quad \text{a.e.} \;\; \sigma \in [0, T]
$$

where $< \cdot >$ denote the inner product in $\mathbb{R}^d$. Fix any interval $[s, t] \subset [0, T]$. It is easy to see that

$$
\int_s^t f(\sigma, x_n(\sigma), u_n(\sigma)) d\sigma \to \int_s^t \dot{x}(\sigma) d\sigma. \tag{3.4}
$$

Now

$$
\begin{aligned}
& \mid \int_s^t (f(\sigma, x(\sigma), u_n(\sigma)) - \dot{x}(\sigma)) d\sigma \mid \\
= & \mid \int_s^t f(\sigma, x_n(\sigma), u_n(\sigma)) - \dot{x}(\sigma)) d\sigma
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_s^t (f(\sigma,x(\sigma),u_n(\sigma)) - f(\sigma,x_n(\sigma),u_n(\sigma))d\sigma \mid \\
&\leq C \sup_\sigma \mid x(\sigma) - x_n(\sigma) \mid \\
&+ \int_s^t (f \mid \sigma, x_n(\sigma), u_n(\sigma)) - \dot{x}(\sigma) \mid ds \\
&\qquad\qquad\qquad \to 0 \qquad (3.5)
\end{aligned}
$$

by (3.3) and (3.4). Next choose $p \in \mathbb{R}^d$. Then by (3.5)

$$< p, \int_s^t (f(\sigma,x(\sigma),u_n(\sigma)) - \dot{x}(\sigma))d\sigma > \qquad \to 0$$

as $n \to \infty$. For any n, however,

$$\int_s^t \sup_{u\in U} < p, f(\sigma,x(\sigma),u) - \dot{x}(\sigma) > d\sigma \ \geq \ < p, \int_s^t (f(\sigma,x(\sigma),u_n(\sigma)) - \dot{x}(\sigma))d\sigma > .$$

Hence

$$\int_s^t \sup_{u\in U} < p, f(\sigma,x(\sigma),u) - \dot{x}(\sigma) > d\sigma \geq 0. \qquad (3.6)$$

Since (3.6) is true for any $[s,t] \subset [0,T]$, we have

$$\sup_{u\in f(t,x(\sigma),U)} < p,u > \quad \geq \quad < p, \dot{x}(\sigma) > \text{a.e.}$$

Let $\mathcal{E}$ be a countable dense subset of $\mathbb{R}^d$. By the above there exists a set $S \subset [0,T]$ of full measure such that

$$\sup_{u\in f(\sigma,x(\sigma),U)} < p,u > \quad \geq \quad < p, \dot{x}(\sigma) > \qquad \forall p \in \mathcal{E}$$

whenever $\sigma \in S$. This implies

$$\dot{x}(\sigma) \in \bar{\text{co}} f(\sigma,x(\sigma),U) = f(\sigma,x(\sigma),U) \qquad (3.7)$$

since by $(A1), (A4), f(\sigma,x(\sigma),U)$ is a closed convex set.
To conclude the proof, we have by $(A3)$,

$$g(x(T)) = \lim_{n\to\infty} g(x_n(T)) = \inf g(y(T)),$$

where the infimum is over all state trajectories $y(\cdot)$, satisfying $y(0) = x$. Now by the foregoing

$$\dot{x}(t) \in f(t,x(t),U) \qquad \text{a.e.}$$

Therefore by a standard measurable selection theorem there exists a control $u(\cdot)$ s.t.

$$\dot{x}(t) = f(t, x(t), u(t)) \qquad \text{a.e.}$$

Hence

$$g(x(T)) = \min_{y(\cdot)} g(y(T)).$$

Thus $u(\cdot)$ is an optimal control. ∎

Remark 3.1. Note that our existence result is heavily based on the assumption $(A4)$. This type of convexity condition is seldom satisfied outside the problems where control enters linearly into the dynamics. It is therefore natural to enquire whether the class of controls can be further extended so as to dispose of this condition. We achieve this by introducing relaxed control problem which we describe below.

Relaxed Control Problem: For any polish space X, $\mathcal{P}(X)$ denotes the space of all probability measures on X. Let $\mathcal{P}(X)$ be endowed with the topology of weak convergence, i.e., the weakest topology that makes $\mathcal{P}(X) \ni \mu \mapsto \int h d\mu \in \mathbb{R}$ continuous for all bounded and continuous map

$$h : X \to \mathbb{R}.$$

Hence

$$\mu_n \to \mu \qquad \text{in } \mathcal{P}(X)$$

if and only if

$$\int h d\mu_n \to \int h d\mu$$

for all $h \in C_b(X)$. For any polish space X, $\mathcal{P}(X)$ is always separable ; $\mathcal{P}(X)$ is compact if and only if X is compact. A relaxed control $\mu(\cdot)$ is a measurable map $\mu : \mathbb{R}_+ \to \mathcal{P}(\mathbb{R}^d)$ such that $\mu(t)(U) = 1$ a.e. A relaxed trajectory $x(\cdot)$ corresponding to the relaxed control $\mu(\cdot)$ is a solution of the equation.

$$\dot{x}(t) = \bar{f}(t, x(t), \mu(t)) \tag{3.8}$$

where

$$\bar{f} : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(U) \to \mathbb{R}^d$$

is defined by

$$\bar{f}(t, x, \mu) = \int_U f(t, x, u)\mu(du). \tag{3.9}$$

Remarks 3.2 If $u(\cdot)$ is a usual control then $\delta_{u(\cdot)}$ is a relaxed control and

$$\bar{f}(t, x, \delta_{u(t)}) = \int f(t, x, u(t)).$$

Therefore the relaxed control setup subsumes the usual one. On the other hand if $f(t,x,U)$ is convex, then under $(A1)$, for any $\mu \in \mathcal{P}(U)$,

$$\bar{f}(t,x,\mu) \in f(t,x,U).$$

Indeed, there exists

$$\mu_n = \sum_{i=1}^{m(n)} \alpha_i^n . \delta_{u_i^n}$$

$$\alpha_i^n \geq 0, \sum_{i=1}^{m(n)} \alpha_i^n = 1, \qquad \text{such that}$$

$$\mu_n \to \mu.$$

Therefore by a standard measurable selection theorem there exist a usual control $u(t)$ such that

$$\bar{f}(t,x,\mu(t)) = f(t,x,u(t)).$$

Hence in this case a relaxed control can be identified with a usual one.
We now show that under the assumptions $(A1)-(A3)$ a relaxed control always exists.

Theorem 3.2 Under $(A1)-(A3)$, an optimal relaxed control exists.

Proof. As before we assume without any loss of generality that $c \equiv 0$. In view of Remark 3.2, it suffixes to show the existence of an optimal usual control under the additional assumption that $f(t,x,U)$ is convex for each t,x, where U is any compact metric space. We take a minimizing of sequence of controls $\{u_n(\cdot)\}$ and $\{x_n(\cdot)\}$ the corresponding trajectories. We can show as before that $\{x_n(\cdot)\}$ is uniformly bounded and uniformly Lipschitz continuous. Therefore $\{x_n(\cdot)\}$ is relatively compact. Let

$$\alpha_n(t) = f(t,x_n(t),u_n(t)).$$

Then $\{\alpha_n(\cdot)\}$ is relatively compact in $L_{2w}[0,T]$. Let $x_n(\cdot) \to x(\cdot)$ in the uniform topology and $\alpha_n(\cdot) \to \alpha(\cdot)$ in $L_{2w}[0,T]$ along the same subsequence. Then

$$\int_0^T h(t)\alpha_n(t)dt \to \int_0^T h(t)\alpha(t)dt$$

for all $h \in L_2[0,T]$. Take $h(t) = I\{0 \leq t \leq s\}$. Then

$$\int_0^s \alpha_n(t)dt \to \int_0^s \alpha(t)dt$$

Since

$$x_n(t) = x + \int_0^t \alpha_n(s)ds,$$

if follows that

$$x(t) = x + \int_0^t \alpha(s)ds.$$

Now, $\alpha_n(\cdot) \to \alpha(\cdot)$ in $L_{2w}[0,T]$. Therefore by Banach - Saks theorem

$$\frac{1}{k}\sum_{i=1}^{k} \alpha_{n_i} \to \alpha(\cdot) \quad \text{in} \quad L_2[0,T].$$

This implies that

$$\frac{1}{k_m}\sum_{i=1}^{k_m} \alpha_{n_i}(t) \to \alpha(t) \qquad \text{a.e. t.}$$

Fix t for which the above holds. Then

$$\begin{aligned} \frac{1}{k_m}\sum_{i=1}^{k_m} \alpha_{n_i}(t) &= \frac{1}{k_m}\sum_{i=1}^{k_m} f(t, x_{n_i}(t), u_{n_i}(t)) \\ &\to \alpha(t). \end{aligned}$$

This implies there exists a $\tilde{u}_i(.)$ such that

$$f(t, x(t), \tilde{u}_i(t)) \to \alpha(t).$$

Since U is compact, $\tilde{u}_i \to u^*$ along a subsequence. Then

$$\alpha(t) = f(t, x(t), u^*) \in f(t, x(t), U) \qquad \text{a.e. } t.$$

On the set of measure zero on which the above fails, redefine $x(t) = f(t, x(t), u_0)$ for a fixed $u_0 \in U$. Therefore there exists a measurable u^* such that $\alpha(t) = f(t, x(t), u^*(t))$. Hence

$$x(t) = x + \int_0^T f(t, x(t), u^*(t))dt.$$

Clearly $u^*(\cdot)$ is an optimal control. ∎

We next state the following theorem.

Theorem 3.3 Assume $(A1) - (A3)$. Let $x^*(\cdot)$ be a relaxed state trajectory. Then given $\varepsilon > 0$ there exists a usual state trajectory $x(\cdot)$ with the same initial condition such that

$$\max_{t\in[0,T]} \mid x^*(\cdot) - x(\cdot) \mid < \varepsilon.$$

As a consequence

$$\begin{aligned} &\min\{g(x(T)) : \quad \text{relaxed state trajectories}\} \\ &= \inf\{g(x(T)) : \quad \text{usual state trajectories}\}. \end{aligned}$$

4 Dynamic Programming

Dynamic programming is a very powerful approach to study dynamic optimization problems. The fundamental principle of dynamic programming, the so called principle of optimality, can be stated quite simply , "If an optimal trajectory is broken into two pieces, then the last piece is itself optimal". In discrete time problems this principle leads to the backward induction algorithm to compute the optimal cost and an optimal control. In continuous time, however, backward induction is not feasible. Nevertheless, this principle of optimality leads to a nonlinear partial differential equation of first order known as Hamilton-Jacobi-Bellman (HJB) equation which in turn determines the optimal cost and an optimal control. To this end we introduce the concept of value function V. Let

$$V : [0, T] \times \mathbb{R}^d \to \mathbb{R}$$

be defined by

$$V(t, x) := \inf_{u(\cdot)} \left[\int_t^T c(s, x(s), u(s))ds + g(x(T)) \right] \tag{4.1}$$

where

$$\begin{aligned} \dot{x}(s) &= f\,(s, x(s), u(s)),\ s \in (t, T)] \\ x(t) &= x \\ u(\cdot) &: [t, T] \to U. \end{aligned}$$

The function V is called the value function and $V(t, x)$ gives the 'minimum cost go' at time t from the state x.

For the sake of simplicity, we assume that the cost functions c and g are bounded below. Since the addition of constants to c and g would not alter an optimal control, we may (and will) assume without any loss of generality that c and g are non-negative functions. Also our aim here is to derive a sufficient condition for optimality. Hence we assume the existence of an optimal control.

Lemma 4.1 Let $x \in \mathbb{R}^d$ and $t \in [0, T], \Delta > 0, t + \Delta \in (0, T]$. Then

$$V(t, x) = \inf_{u(.)} \left[\int_t^{t+\Delta} c(s, x(s), u(s))ds + V(t + \Delta, x(t + \Delta)) \right] \tag{4.2}$$

where

$$\dot{x}(s) = f(s, x(s), u(s)), \quad s \in (t, T]\,, x(t) = x$$

and

$$u(\cdot) : [t, T] \to U.$$

Proof. For any measurable function $u(\cdot) : [t,T] \to U$, we have by the definition of V

$$V(t,x) \le \int_t^{t+\Delta} c(s,x(s),u(s))ds + \int_{t+\Delta}^T c(s,x(s),u(s))ds + g(x(T)).$$

Taking infimum in the second and third terms in the above, we get

$$V(t,x) \le \int_t^{t+\Delta} c(s,x(s),u(s))ds + V(t+\Delta, x(t+\Delta)).$$

Therefore

$$V(t,x) \le \inf_{u(.)} \left[\int_t^{t+\Delta} c(s,x(s),u(s))ds + V(t+\Delta, x(t+\Delta)) \right]. \tag{4.3}$$

To get the reverse inequality, let $\varepsilon > 0$ be arbitrary. Then by the definition of $V(t,x)$, there exists a control $u_\varepsilon(\cdot)$ such that

$$V(t,x) \ge \int_t^{t+\Delta} c(s,x_\varepsilon(s),u_\varepsilon(s))ds + \int_{t+\Delta}^T c(s,x_\varepsilon(s),u_\varepsilon(s))ds + g(x_\varepsilon(T)) - \varepsilon, \tag{4.4}$$

where $x_\varepsilon(\cdot)$ is the state under $u_\varepsilon(\cdot)$ with $x_\varepsilon(t) = x$. Taking infimum in the second and third terms in (4.4) we obtain

$$V(t,x) \ge \inf_{u(\cdot)} \left[\int_t^{t+\Delta} c(s,x(s),u(s))ds + V(t+\Delta, x(t+\Delta)) \right] - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary

$$V(t,x) \ge \inf_{u(\cdot)} \left[\int_t^{t+\Delta} c(s,x(s),u(s))ds + V(t+\Delta, x(t+\Delta) \right]. \tag{4.5}$$

The desired result (4.2) follows from (4.3) and (4.5). ∎

Remark 4.1. The equation (4.2) is the integral from of the dynamic programming equation. The Hamilton-Jacobi-Bellman (HJB) equation is usually derived from (4.2).

Exercise 4.1 Assume $c \equiv 0$. Let $u(\cdot)$ be a control and $x(\cdot)$ the corresponding state trajectory. Show that $V(t,x(t))$ is a nondecreasing function of t.

Exercise 4.2 Assume $c \equiv 0$. Let $u^*(\cdot)$ be an optimal control and $x^*(\cdot)$ the corresponding state trajectory. Show that $V(t,x^*(t))$ is a constant, i.e., it is independent of t.

Exercise 4.3 Show that V is Lipschitz continuous, i.e., there exists a constant $K > 0$ such that for $t_1, t_2 \in [0,T]$, $x_1, x_2 \in \mathbb{R}^d$

$$| \, V(t_1,x_1) - V(t_2,x_2) \, | \le K \left(| \, t_1 - t_2 \, | + | \, x_1 - x_2 \, | \right).$$

(Hint. Use the Lipschitz continuity properties of c, f and g.)

Remark 4.2 It follows from Exercise 4.3 that V is differentiable almost everywhere. Under the additional assumption that $V \in C^{1,1}([0,T] \times \mathbb{R}^d)$ we can establish the following result.

Theorem 4.1 Let $V \in C^{1,1}([0,T] \times \mathbb{R}^d)$. Assume that an optimal control exists. Then V satisfies the following nonlinear partial differential equation known as HJB equation

$$\left.\begin{array}{lcl} V_t(t,x) & + & \inf_{u \in U} [\nabla_x V(t,x).f(t,x,u) + c(t,x,u)] = 0 \\ \\ V(T,x) & = & g(x) \qquad . \end{array}\right\} \tag{4.6}$$

Proof. Let $t \in [0,T], x \in \mathbb{R}^d$, $\Delta > 0, t + \Delta \in (0,T]$. From Lemma 4.1

$$V(t,x) = \inf_{u(.)} \left[\int_t^{t+\Delta} c(s,x(s),u(s))ds + V(t+\Delta, x(t+\Delta)) \right].$$

Let $u(\cdot)$ be a control with $u(t) = u_0$ and $u(\cdot)$ continuous at t. Then

$$0 \leq \int_t^{t+\Delta} c(s,x(s),u(s))ds + V(t+\Delta, x(t+\Delta)) - V(t,x).$$

Dividing by Δ and letting $\Delta \to 0$, we get

$$V_t(t,x) + \nabla_x V(t,x).f(t,x,u_0) + c(t,x,u_0) \geq 0 \tag{4.7}$$

for all $u_0 \in U$. To obtain the reverse inequality let $u^*(\cdot)$ be an optimal control with $u^*(t) = u^*$ and $x^*(\cdot)$ the corresponding state trajectory with $x^*(t) = x$. Assume that $c \equiv 0$. Then by Exercise 4.2, $V(s, x^*(s))$ is independent of s. Therefore

$$\frac{d}{ds} V(s, x^*(s)) = 0.$$

Thus

$$V_t(t,x) + \nabla_x V(t,x).f(t,x,u^*) = 0 \tag{4.8}$$

If $c \neq 0$, then (4.8) takes the form

$$V_t(t,x) + \nabla_x V(t,x).f(t,x,u^*) + c(t,x,u^*) = 0. \tag{4.9}$$

The desired result the follows from (4.7) and (4.9). ■

Remark 4.3 Define the function

$$H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$$

by

$$H(t,x,u,p) = p.f(t,x,u) + c(t,x,u). \tag{4.10}$$

The function H is called the (pseudo) Hamiltonian for the system. In terms of H the equation (4.6) becomes

$$\left.\begin{array}{rcl} V_t(t,x) & + & \inf_{u\in U} H(t,x,u,\nabla_x V(t,x)) = 0 \\ \\ V(T,x) & = & g(x) \qquad . \end{array}\right\} \tag{4.11}$$

The value function $V(t,x)$, however, need not be continuously differentiable as the following example shows (Zhou (1993)).

Example 4.1 Consider the following optimal control problem : minimize $-x(1)$ subject to :

$$\begin{array}{rcl} \dot{x}(t) & = & x(t)u(t),\ t \in (0,1] \\ \\ x(0) & = & x \\ \\ u(\cdot) & : & [0,1] \to [0,1]\,. \end{array}$$

It is easily seen that the value function $V(t,x)$ is given by

$$V(t,x) = \begin{cases} -xe^{1-t} & , \quad x > 0 \\ \\ -x & , \quad x \le 0. \end{cases}$$

Indeed the existence of a $C^{1,1}([0,T]) \times \mathbb{R}^d)$ solution of (4.11) is more of an exception than a rule. Nevertheless, the existence of a $C^{1,1}([0,T]) \times \mathbb{R}^d)$ solution of (4.11) leads to the following verification theorem.

Theorem 4.2 Let $W \in C^{1,1}([0,T] \times \mathbb{R}^d)$ be a solution of (4.11). Then
(a) W satisfies

$$W(t,x) \le \int_t^T c(s,x(s),u(s))ds + g(x(T))$$

for any control $u(\cdot)$ where

$$\begin{array}{rcl} \dot{x}(s) & = & f(s,x(s),u(s))ds + g(x(T)) \\ \\ x(t) & = & x. \end{array}$$

Thus $W(t,s) \ \le \ V(t,x)$.
(b) Suppose $u^*(\cdot)$ is a control and $x^*(\cdot)$ the corresponding state trajectory. If

$$W_t(t,x^*(t)) + H(t,x^*(t),\nabla_x W(t,x^*(t))) = 0 \qquad \text{a.e.} \tag{4.12}$$

then $u^*(\cdot)$ is an optimal control and $W(t,x) = V(t,x)$.

Proof. (a) We have

$$\frac{d}{ds}W(t,x(s)) \ = \ W_s(t,x(s)) + \nabla_x W(s,x(s)).f(s,x(s),u(s))$$

$$\geq \quad -c(s, x(s), u(s)).$$

Therefore

$$g(x(T)) - W(t, x) \geq - \int_t^T c(s, x(s), u(s))ds.$$

Thus

$$W(t, x) \leq \int_t^T c(s, x(s, u(s))ds + g(x(T)).$$

(b) Arguing as above, we have

$$W(t, x) = \int_t^T c(s, x^*(s), u^*(s))ds + g(x^*(T)).$$

It thus follows that

$$W(t, x) = V(t, x).$$

and hence $u^*(.)$ is optimal. ■

Remark 4.4 It follows from the above theorem that if (4.11) has a solution $W \in C^{1,1}([0, T]) \times \mathbb{R}^d)$, and furthermore if (4.12) is satisfied then the solution W equals the value function V. In other words, under these circumstances the value function V is the unique $C^{1,1}([0, T] \times \mathbb{R}^d)$ solution of (4.11). But the value function need not be $C^{1,1}$ as we have seen in Example 4.1. Moreover, there are practical difficulties in verifying (4.12) with W replaced by the value function V. In practice Theorem 4.2 is used via a 'feedback' control which we describe now. If there exists a 'nice' function

$$\bar{u} : [0, T] \times \mathbb{R}^d \to U$$

satisfying

$$\begin{aligned} & c\,(t, x, \bar{u}(x, t)) + \nabla_x V(t, x).f(t, x, \bar{u}(t, x)) \\ & = \min_{u \in U} [c(t, x, u) + \nabla_x V(t, x).f(t, x, u)] \end{aligned} \tag{4.13}$$

such that the 'closed-loop' equation

$$\left.\begin{aligned} \dot{x}(t) &= f(t, x(t), \bar{u}(t, x)) \\ x(0) &= x \end{aligned}\right\} \tag{4.14}$$

has a unique solution, then the feedback control given by

$$u^*(t) = \bar{u}(t, x(t)) \tag{4.15}$$

is optimal. Stringent conditions on f and c are, however, needed to ensure the validity of (4.13) and (4.14). Nevertheless, when these conditions are satisfied, (4.13) - (4.15) offer a very useful way of finding an optimal control. To illustrate this point we describe the celebrated LQ problem which is an optimal control problem involving a linear system with quadratic cost.

Example 4.2 Minimize

$$\int_0^T [a(t)x^2(t) + b(t)u^2(t)]dt + cx^2(T)$$

subject to

$$\dot{x}(t) = \alpha(t)x(t) + \beta(t)u(t),\ x(0) = x \in \mathbb{R}$$

where

$$a(t) \geq 0, b(t) > 0, c \geq 0, u(t) \in \mathbb{R}.$$

The HJB equation (4.11) for this problem is given by

$$\left.\begin{array}{c} \frac{\partial V}{\partial t} = -\min_{u \in I\!R} \left[a(t)x^2 + b(t)u^2 + \dfrac{\partial V}{\partial x}(\alpha(t)x + \beta(t)u) \right] \\ V(T,x) = cx^2. \end{array}\right\} \tag{4.16}$$

The minimum on the right side of (4.16) is achieved at

$$u^*(t) = -\frac{\beta(t)}{2b(t)}\frac{\partial V}{\partial x}. \tag{4.17}$$

Substituting (4.17) into (4.16) we get

$$-\frac{\partial V}{\partial t} = a(t)x^2 - \frac{\beta^2(t)}{4b(t)}\left(\frac{\partial V}{\partial x}\right)^2 + \alpha(t)x\frac{\partial V}{\partial x}. \tag{4.18}$$

We look for a trial solution

$$V(t,x) = p(t)x^2 + q(t).$$

Then

$$\frac{\partial V}{\partial x} = 2p(t)x, \qquad \frac{\partial V}{\partial t} = \dot{p}(t)x^2 + \dot{q}(t).$$

Substituting these into (4.18) and simplifying we obtain

$$-\dot{p}(t)x^2 - \dot{q}(t) = \left[a(t) - \frac{\beta^2(t)p^2(t)}{b(t)} + 2\alpha(t)p(t)\right]x^2$$

Thus

$$\left.\begin{array}{rcl} \dot{p}(t) &=& \frac{\beta^2(t)p^2(t)}{b(t)} - a(t) - 2\alpha(t)p(t) \\ \\ p(T) &=& c. \end{array}\right\} \tag{4.19}$$

Also, $\dot{q}(t) = 0$ and $p(T)x^2 + q(T) = cx^2$, which implies that $q(T) = 0$. Hence $q(t) \equiv 0$. Therefore

$$V(t,x) = p(t)x^2$$

is the value function, where $p(\cdot)$ is the solution of the Riccati equation (4.19). The optimal feedback control is given by

$$\bar{u}(t,x) = -\frac{\beta(t)}{b(t)}p(t)x.$$

Exercise 4.4 Consider the general LQ problem: minimize

$$\int_0^T [x'(t)Q(t)x(t) + u'(t)R(t)u(t)]\,dt + x'(T)Sx(T)$$

subject to :

$$\dot{x}(t) = A(t)x(t) + B(t)u(t),\ x(0) = x \in \mathbb{R}^d$$

$u(t) \in \mathbb{R}^m, Q(t)$ and S are non negative definite $d \times d$ matrix, and $R(t)$ is an $m \times m$ positive definite matrix, $A(t)$ is a $d \times d$ matrix, and $B(t)$ is a $d \times m$ matrix. Find the value function and the optimal feedback control for this problem.

Exercise 4.5 Calculate the feedback control that minimizes

$$\int_0^T \left[x_1(t) - \alpha)^2 + u^2(t)\right] dt$$

for the system

$$\dot{x}_1(t) = x_2, \qquad \dot{x}_2(t) = u(t)$$

$$u(t) \in \mathbb{R}, \qquad x_1(0) = x_1, \qquad x_2(0) = x_2.$$

Remark 4.5 We have seen in Example 4.1 that the value function V need not be differentiable. This makes the applicability of Theorem 4.2 rather limited. On the other hand from Exercise 4.3. it follows that V is differentiable almost everywhere. One can then modify the proof of Theorem 4.1 to show that V satisfies (4.11) almost everywhere. We can also modify the proof of Theorem 4.2 to show that if there exists a control $u^*(\cdot)$ such that

$$\begin{aligned} &H(t, x^*(t), u^*(t), \nabla_x V(t, x^*(t)) \\ &= \inf_{u \in U} H(t, x^*(t), \nabla_x V(t, x^*(t))) \qquad \text{a.e.t.} \end{aligned}$$

then $u^*(\cdot)$ is optimal. We cannot, however, show that V is the unique solution of (4.11) in the class of functions which are differentiable almost everywhere. Moreover, the above cannot be directly used to get a feedback control. These difficulties can be overcome by introducing a new solution concept for the equation of the type (4.11), namely the viscosity solution (Fleming and Soner (1991)). The verification theorem (Theorem 4.2) can also be generalized in terms of the viscosity solution (Zhou (1993)). We leave these topics for further reading.

5 Conclusions

In this note we have formulated the optimal control problem and presented the existence theory. We have also studied the dynamic programming approach to optimal control problems. Here we have not studied Pontryagin's maximum principle which provides a very powerful method in finding optimal controls. For these issues and many more information on optimal control we refer to Fleming and Rishel (1975), Fleming and Soner (1991), and Clarke (1983).

References

[1] F. H. Clarks (1983), Optimization and Nonsmooth Analysis, Wiley.

[2] W. H. Fleming and R. W. Rishel (1975), Deterministic and Stochastic Optimal Control, Springer-Verlag.

[3] W.H. Fleming and H.M. Soner (1991), Controlled Markov Processes and Viscosity Solution, Springer-Verlag.

[4] S.Y. Zhou (1993), Verification theorems withing the famework of viscosity solutions, J. Math. Anal. Appl. 177, 208-225.