

# HOPF–LAX FORMULA FOR VARIATIONAL PROBLEMS WITH NON–CONSTANT DISCOUNT

JUAN PABLO RINCÓN–ZAPATERO

ABSTRACT. We provide a Hopf–Lax formula for variational problems with non–constant discount and deduce a dynamic programming equation. We also study some regularity properties of the value function.

## 1. INTRODUCTION

We establish a Hopf–Lax formula for the Cauchy problem

$$(1) \quad \begin{cases} -v_t(x, t) + f(-v_x) + \rho(t)v(x, t) = 0, & \text{in } \mathbb{R}^n \times (0, T); \\ v = g, & \text{on } \mathbb{R}^n \times \{t = T\}, \end{cases}$$

involving a Hamilton–Jacobi equation with a linear dissipation term,  $\rho(t)v(x, t)$ , and a terminal condition at time  $t = T$ . The function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is assumed to be convex and of class  $C^2$ ,  $\rho : [0, T] \rightarrow (0, 1]$  is continuous, and  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is globally Lipschitz. The formula is

$$v(x, t) = \min_{p \in \mathbb{R}^n} \left\{ \int_t^T d_t(s) \ell(\iota(d_t^{-1}(s)p)) ds + d_t(T)g \left( x + \int_t^T \iota(d_t^{-1}(s)p) ds \right) \right\},$$

where  $\ell$  is the convex conjugate of  $f$ ,  $\iota = (\nabla \ell)^{-1}$ , and  $d_t(s) = \exp(-\int_t^s \rho(r) dr)$ . The formula represents a Lipschitz solution that satisfies the Cauchy problem almost everywhere.

The classical Hopf–Lax formula applies to the case  $\rho \equiv 0$ , and was given by Lax in [6], for  $n = 1$ . It was extended later to general  $n$  by Hopf in [5]. Further generalizations have maintained  $\rho = 0$  but have considered functions  $f(t, -v_x)$  also depending on time, [9]; or functions  $f(v, -v_x)$  depending on both  $v$  and  $v_x$ , [2], with some additional requirements. The case we analyze in this paper is not covered in any of these previous works.

Actually, we find a Hopf–Lax formula that applies to more general Hamilton–Jacobi equations associated to the calculus of variations problems with variable discount.

These problems arise quite naturally in models of economics. Consider for instance the following problem: an agent optimally chooses a consumption path of a given good, with the aim of maximizing his/her satisfaction. This is measured by an utility function of

---

*Date:* June 24, 2009.

I am grateful to two anonymous referees for useful comments. I acknowledge the support of the Spanish Ministry of Ciencia e Innovación through grant ECO2008–02358.

consumption,  $\ell(u)$ , along a given time interval,  $[0, T]$ . It is customary in the literature to postulate concavity in the preferences of the agent, and to suppose that he/she is impatient, in the sense that the value of the utility attained today is higher than the utility attained tomorrow. This is the meaning of introducing a discount factor or impatience rate in the preferences of the agent.

Empirical studies suggest that people are more impatient about choices in the short run than in the long run, implying that the discount rate applied to current choices is higher than the one applied to far-in-the future choices. Thus, the discount factor should be taken to be non-constant. Several papers have considered the non-constant discount case: see e.g. [1], [4] or [7]<sup>1</sup>. In [1], the optimal growth model with time-varying discount is considered, for a particular class of utility functions. A general problem, with infinite horizon, is analyzed in [4], whereas [7] considers the finite horizon case with fixed or variable terminal time. The last two papers use discretization and passage to the limit to find a Hamilton–Jacobi equation that involves not only the unknown value function, but also a non-local term involving integration throughout the unknown optimal solution. We provide conditions so that the Hamilton–Jacobi equation involves only the derivatives of the value function and find the dynamic programming equation by direct methods.

Given the significance of the non-constant discount preference rate in economics, it is of interest to analyze in more detail this type of variational problems. First, deriving a Hopf–Lax formula for the solution of the variational problem (Section 3); second, establishing a modified dynamic programming equation, more amenable than the one found in previous papers (Section 4); and third, studying the regularity of the value function (Section 5).

## 2. VARIATIONAL PROBLEM WITH DISCOUNT

We follow the presentation in [3]. Let the value function

$$(2) \quad v(x, t) = \inf_{y \in AC_{x,t}} \left\{ \int_t^T d_t(s) \ell(\dot{y}(s)) ds + d_t(T) g(y(T)) \right\},$$

where

$$AC_{x,t} = \{y : [t, T] \longrightarrow \mathbb{R}^n : y = y(s) \text{ absolutely continuous, } y(t) = x\}.$$

A typical element of this set will be called an arc. We will impose the following conditions.

**A1:**  $\ell : \mathbb{R}^n \longrightarrow \mathbb{R}$  is  $C^2$ , strictly convex, and  $\lim_{|u| \rightarrow \infty} \frac{\ell(u)}{|u|} = \infty$ ;

**A2:**  $g$  is globally Lipschitz in  $\mathbb{R}^n$ ;

**A3:**  $d : [0, T] \times [0, T] \longrightarrow (a, 1]$ , with  $a > 0$ , is Lipschitz continuous with  $d_t(t) = 1$  for each  $t$ .

---

<sup>1</sup>We consider only papers on continuous time.

A straightforward interpretation of (2) has been done in the Introduction: a single agent, with time-varying preference rate, chooses optimally over time. In another reading, there is a continuum of agents, each one labelled by  $t \in [0, T]$ ; each agent (or generation)  $t$  applies a possibly different discount factor,  $d_t$ , in the calculation of the utility flow from  $t$  onwards. At time  $T$  the optimization process finishes and agent  $T$  derives utility  $g(y(T))$  (or “scrap value”). The aim of each generation is to maximize the total discounted utility. In this process, the  $t$ -generation is not so much concerned with the consumption of the future generations as it is with respect its own consumption.

A common specification of  $d_t(s)$  is

$$(3) \quad d_t(s) = \exp\left(-\int_t^s \rho(r) dr\right),$$

where  $\rho \in L^\infty([0, T])$ . In this case  $d_t(s)$  is Lipschitz in  $(t, s)$ , which is the present value at time  $t$  of one unit of utility at time  $s \geq t$ . The rate of discount is  $\rho$ , and most often it is considered constant. Other popular discount factors are those that depend only on the elapsed time,  $d_t(s) = \theta(s - t)$  for  $s \geq t$ , through a scalar function  $\theta$ , with  $\theta(0) = 1$ . As will be seen in Section 4, the shape of the discount factor has a major effect in the structure of the dynamic programming equation.

Let us define  $\iota = (\nabla\ell)^{-1}$ , the inverse of  $\nabla\ell$ . Notice that by A1, both  $\nabla\ell$  and  $\iota$  are continuous, and surjective. We also consider  $\ell^*(p) = \sup_{u \in \mathbb{R}^n} \{p \cdot u - \ell(u)\}$ , the Legendre transform of  $\ell$ . Finally, let the  $t$ -Hamiltonian

$$H_t(s, u, p) = p \cdot u - d_t(s)\ell(u).$$

Throughout the paper,  $\nabla$  denotes the gradient of a real function, and  $\nabla^2$  the Hessian Matrix. For a vector function,  $\nabla$  denotes the Jacobian matrix.

### 3. HOPF–LAX FORMULA

A Hopf–Lax formula describes an infinite dimensional variational problem as a finite dimensional one. In the present case, the formula is a little more involved than in the non-discounted case, due to the non-autonomous term  $d_t(s)$ . Notice also that the problem at hand is different from the one with a non-autonomous  $\ell(s, \dot{y}(s))$ , because the current date  $t$  enters into the definition.

Given  $t \in [0, T]$ ,  $t \leq s \leq T$ ,  $x, \alpha \in \mathbb{R}^n$ , consider

$$\begin{aligned} U_{t,\alpha}(s) &= \iota(d_t^{-1}(s)\nabla\ell(\alpha)), \quad (d_t^{-1} = 1/d_t) \\ Y_{x,t,\alpha}(s) &= x + \int_t^s \iota(d_t^{-1}(r)\nabla\ell(\alpha)) dr, \\ V(x, t, \alpha) &= \int_t^T d_t(s)\ell(U_{t,\alpha}(s)) ds + d_t(T)g(Y_{x,t,\alpha}(T)). \end{aligned}$$

Notice that  $Y_{x,t,\alpha}(s)$  is absolutely continuous and  $Y_{x,t,\alpha}(t) = x$ , thus it is an admissible arc, i.e. it belongs to  $AC_{x,t}$ . Observe also that  $\dot{Y}_{x,t,\alpha}(s) = U_{t,\alpha}(s)$ .

We establish the following lemma to facilitate posterior quotation. It is a consequence of assumption A1.

**Lemma 3.1.** *For  $x \in \mathbb{R}^n$ ,  $t \in [0, T]$ ,  $s \geq t$ , the mappings  $\alpha \mapsto U_{t,\alpha}(s)$ ,  $\alpha \mapsto Y_{x,t,\alpha}(s)$  are of class  $C^1$  and surjective.*

**Theorem 3.1.** (Hopf–Lax formula with discount). *If  $x \in \mathbb{R}^n$  and  $0 \leq t < T$ , then the value function  $v = v(x, t)$  of the minimization problem (2) is given by*

$$(4) \quad v(x, t) = \min_{p \in \mathbb{R}^n} \left\{ \int_t^T d_t(s)\ell(\iota(d_t^{-1}(s)p)) ds + d_t(T)g\left(x + \int_t^T \iota(d_t^{-1}(s)p) ds\right) \right\}.$$

*Proof.* 1. For any  $\alpha \in \mathbb{R}^n$

$$v(x, t) \leq \int_t^T d_t(s)\ell(\dot{Y}_{x,t,\alpha}(s)) ds + d_t(T)g(Y_{x,t,\alpha}(T)) = V(x, t, \alpha),$$

and so

$$v(x, t) \leq \inf_{\alpha \in \mathbb{R}^n} V(x, t, \alpha).$$

2. On the other hand, for an arbitrary arc  $y(s)$ ,  $t \leq s \leq T$ , with  $y(t) = x$ , let  $\bar{\alpha}$  be such that

$$Y_{x,t,\bar{\alpha}}(T) = y(T).$$

This is possible by Lemma 3.1. For each  $t, s, p$ , the Hamiltonian  $H_t(\cdot, u, \cdot)$  is concave, thus for any  $\alpha$

$$(5) \quad H_t(s, U_{t,\alpha}(s), \nabla\ell(\alpha)) \geq H_t(s, \dot{y}(s), \nabla\ell(\alpha)),$$

since

$$\left. \frac{\partial H_t(s, u, \nabla\ell(\alpha))}{\partial u} \right|_{u=U_{t,\alpha}(s)} = 0.$$

Let  $\alpha = \bar{\alpha}$  defined above. Integrating (5) between  $t$  and  $T$  and rearranging terms we get

$$\begin{aligned} \int_t^T d_t(s)\ell(U_{t,\bar{\alpha}}(s)) ds &\leq \int_t^T d_t(s)\ell(\dot{y}(s)) ds + \nabla\ell(\bar{\alpha}) \int_t^T (U_{t,\bar{\alpha}}(s) - \dot{y}(s)) ds \\ &= \int_t^T d_t(s)\ell(\dot{y}(s)) ds + \nabla\ell(\bar{\alpha}) \int_t^T (\dot{Y}_{x,t,\bar{\alpha}}(s) - \dot{y}(s)) ds \\ &= \int_t^T d_t(s)\ell(\dot{y}(s)) ds, \end{aligned}$$

because  $Y_{x,t,\bar{\alpha}}(t) = x = y(t)$  and  $Y_{x,t,\bar{\alpha}}(T) = y(T)$ . Adding  $d_t(T)g(Y_{x,t,\bar{\alpha}}(T)) = d_t(T)g(y(T))$  to both terms of the above inequality we get that, for any arc  $y(s)$ , there exist some  $\bar{\alpha}$  such that

$$V(x, t, \bar{\alpha}) \leq \int_t^T d_t(s)\ell(\dot{y}(s)) ds + d_t(T)g(y(T)).$$

Thus  $\inf_{\alpha \in \mathbb{R}^n} V(x, t, \alpha) \leq v(x, t)$ . Hence,  $\inf_{\alpha \in \mathbb{R}^n} V(x, t, \alpha) = v(x, t)$ . Finally, observe that minimization with respect to  $\alpha$  is equivalent of minimization with respect to  $p = \nabla\ell(\alpha)$ .

3. The infimum is in fact attained, since the function  $V(\cdot, \cdot, \alpha)$  is continuous and inf-compact. Indeed,  $\lim_{|\alpha| \rightarrow \infty} |\alpha|^{-1}V(x, t, \alpha) = \infty$  due to the assumptions A1–A3 and Lemma 3.1.  $\square$

Define  $A(x, t) = \operatorname{argmin}_{\alpha \in \mathbb{R}^n} V(x, t, \alpha)$ . Since  $\lim_{|\alpha| \rightarrow \infty} |\alpha|^{-1}V(x, t, \alpha) = \infty$ ,  $A$  is compact valued and an upper semicontinuous correspondence.

The following corollary is along the lines of the above proof.

**Corollary 3.1.** *If  $x \in \mathbb{R}^n$  and  $0 \leq t < T$ , then for any selection  $\alpha(x, t) \in A(x, t)$ , the arc  $Y_{x,t,\alpha(x,t)}(s)$  is a solution of problem (2).*

**Remark 3.1.** *When  $d_t(s) = 1$  for all  $0 \leq t \leq s \leq T$ , (4) reduces to the classical Hopf–Lax formula*

$$v(x, t) = \min_{\alpha \in \mathbb{R}^n} \left\{ (T - t)\ell\left(\frac{\alpha - x}{T - t}\right) + g(\alpha) \right\},$$

after substituting  $\alpha = x + (T - t)\iota(p)$ .

For a locally Lipschitz function  $f$ ,  $\operatorname{lip}(f)$  will denote the Lipschitz parameter of  $f$  in a given compact set  $K$ , and  $\operatorname{bound}(f)$  will denote a bound of  $|f|$  in that set. Notice that under our assumptions  $f = \iota, d, d^{-1}$  are locally Lipschitz.

**Theorem 3.2.** (Lipschitz continuity). *The value function  $v$  is locally Lipschitz continuous in  $\mathbb{R}^n \times [0, T]$  and*

$$v = g \quad \text{on } \mathbb{R}^n \times \{t = T\}.$$

*Proof.* Let  $x, \hat{x} \in K \subseteq \mathbb{R}^n$  with  $K$  compact,  $t, \hat{t} \in [0, T)$  and  $\alpha \in \mathbb{R}^n$ . Let  $\hat{t} \wedge t = \min\{\hat{t}, t\}$  and  $\hat{t} \vee t = \max\{\hat{t}, t\}$ .

1. Let  $s \in [0, T)$  and  $\alpha \in \mathbb{R}^n$ . Let us proceed to establish some Lipschitz estimates.

$$\begin{aligned} |U_{\hat{t}, \alpha}(s) - U_{t, \alpha}(s)| &\leq \text{lip}(\iota) |\nabla \ell(\alpha)| |d_{\hat{t}}^{-1}(s) - d_t^{-1}(s)| \\ &\leq \text{lip}(\iota) |\nabla \ell(\alpha)| \text{lip}(d^{-1}) |\hat{t} - t| = C_1 |\hat{t} - t|. \end{aligned}$$

$$\begin{aligned} (6) \quad &\int_{\hat{t}}^T d_{\hat{t}}(s) \ell(U_{\hat{t}, \alpha}(s)) ds - \int_t^T d_t(s) \ell(U_{t, \alpha}(s)) ds \\ &\leq \int_{\hat{t} \vee t}^T d_{\hat{t} \vee t}(s) |\ell(U_{\hat{t}, \alpha}(s)) - \ell(U_{t, \alpha}(s))| ds + \int_{\hat{t} \wedge t}^{\hat{t} \vee t} |d_{\hat{t} \wedge t}(s) - d_{\hat{t} \vee t}(s)| |\ell(U_{\hat{t} \wedge t, \alpha}(s))| ds \\ &\leq \text{bound}(d) \text{lip}(\ell) C_1 |\hat{t} - t| + T \text{lip}(d) \text{bound}(\ell) |\hat{t} - t| \\ &= C_2 |\hat{t} - t|. \end{aligned}$$

$$\begin{aligned} |Y_{\hat{x}, \hat{t}, \alpha}(T) - Y_{x, t, \alpha}(T)| &\leq |\hat{x} - x| + \left| \int_{\hat{t}}^T U_{\hat{t}, \alpha}(s) ds - \int_t^T U_{t, \alpha}(s) ds \right| \\ &\leq |\hat{x} - x| + \int_{\hat{t} \vee t}^T |U_{\hat{t}, \alpha}(s) - U_{t, \alpha}(s)| ds + \int_{\hat{t} \wedge t}^{\hat{t} \vee t} |U_{\hat{t} \wedge t, \alpha}(s)| ds \\ &\leq |\hat{x} - x| + TC_1 |\hat{t} - t| + \text{bound}(U) |\hat{t} - t| \\ &= |\hat{x} - x| + C_3 |\hat{t} - t|. \end{aligned}$$

$$\begin{aligned} (7) \quad |d_{\hat{t}}(T)g(Y_{\hat{x}, \hat{t}, \alpha}(T)) - d_t(T)g(Y_{x, t, \alpha}(T))| &\leq |d_{\hat{t}}(T) - d_t(T)| |g(Y_{\hat{x}, \hat{t}, \alpha}(T))| \\ &\quad + d_t(T) |g(Y_{\hat{x}, \hat{t}, \alpha}(T)) - g(Y_{x, t, \alpha}(T))| \\ &\leq \text{lip}(d) \text{bound}(g) |\hat{t} - t| \\ &\quad + \text{lip}(g) \text{bound}(d) (|\hat{x} - x| + C_3 |\hat{t} - t|) \\ &\leq C_4 (|\hat{x} - x| + |\hat{t} - t|). \end{aligned}$$

2. Choose  $\alpha \in A(x, t)$ . Then, by definition of  $v$  and estimates (6) and (7)

$$\begin{aligned} v(\hat{x}, \hat{t}) - v(x, t) &\leq V(\hat{x}, \hat{t}, \alpha) - V(x, t, \alpha) \\ &= \int_{\hat{t}}^T d_{\hat{t}}(s) \ell(U_{\hat{t}, \alpha}(s)) ds - \int_t^T d_t(s) \ell(U_{t, \alpha}(s)) ds \\ &\quad + d_{\hat{t}}(T)g(Y_{\hat{x}, \hat{t}, \alpha}(T)) - d_t(T)g(Y_{x, t, \alpha}(T)) \\ &\leq C_2 |\hat{t} - t| + C_4 (|\hat{x} - x| + |\hat{t} - t|). \end{aligned}$$

Reversing the role of  $(\hat{x}, \hat{t})$  and  $(x, t)$  we get the desired Lipschitz property.

3. Now, let  $x \in \mathbb{R}^n$ ,  $t < T$  and define  $\delta_t = \int_t^T d_t(s) ds$ . Choose  $\alpha \in \mathbb{R}^n$  such that  $\int_t^T U_{t,\alpha}(s) ds = 0$ ; this is possible by virtue of Lemma 3.1. Let  $b = \max_{s \in [t, T]} |U_{t,\alpha}(s)|$ , and let  $\text{bound}(\ell)$  be a bound of  $|\ell|$  in  $[-b, b]$ . Then,

$$(8) \quad v(x, t) \leq \int_t^T d_t(s) \ell(U_{t,\alpha}(s)) ds + d_t(T)g(x) \leq \text{bound}(\ell)\delta_t + d_t(T)g(x).$$

Moreover,

$$\begin{aligned} v(x, t) &\geq d_t(T)g(x) + \min_{\alpha \in \mathbb{R}^n} \left\{ -\text{lip}(g)d_t(T) \left| \int_t^T U_{t,\alpha}(s) ds \right| + \int_t^T d_t(s) \ell(U_{t,\alpha}(s)) ds \right\} \\ &\geq d_t(T)g(x) + d_t(T)\delta_t \min_{\alpha \in \mathbb{R}^n} \left\{ -\text{lip}(g)\delta_t^{-1} \left| \int_t^T U_{t,\alpha}(s) ds \right| + \ell \left( \delta_t^{-1} \int_t^T d_t(s) U_{t,\alpha}(s) ds \right) \right\}, \end{aligned}$$

where the first inequality holds since  $g$  is Lipschitz, and the second one is simply Jensen's inequality. Now, notice that for any  $\alpha \in \mathbb{R}^n$

$$\frac{\int_t^T U_{t,\alpha}(s) ds}{\int_t^T d_t(s) U_{t,\alpha}(s) ds} \rightarrow 1 \quad \text{as } t \rightarrow T^- \quad (\text{componentwise})$$

thus, for every  $t$  close enough to  $T$ , there exists  $\epsilon > 0$  such that

$$\begin{aligned} v(x, t) &\geq d_t(T)g(x) + d_t(T)\delta_t \min_{\alpha \in \mathbb{R}^n} \left\{ -\text{lip}(g)(1 + \epsilon)\delta_t^{-1} \left| \int_t^T d_t(s) U_{t,\alpha}(s) ds \right| \right. \\ &\quad \left. + \ell \left( \delta_t^{-1} \int_t^T d_t(s) U_{t,\alpha}(s) ds \right) \right\} \\ &= d_t(T)g(x) - d_t(T)\delta_t \max_{z \in B} \max_{\alpha \in \mathbb{R}^n} \left\{ z\delta_t^{-1} \int_t^T d_t(s) U_{t,\alpha}(s) ds \right. \\ &\quad \left. - \ell \left( \delta_t^{-1} \int_t^T d_t(s) U_{t,\alpha}(s) ds \right) \right\}, \end{aligned}$$

where  $B = [-\text{lip}(g)(1 + \epsilon), \text{lip}(g)(1 + \epsilon)]$ . Then

$$(9) \quad v(x, t) \geq d_t(T)g(x) - d_t(T)\delta_t \max_{z \in [-\text{lip}(\ell)(1+\epsilon), \text{lip}(\ell)(1+\epsilon)]} \ell^*(z),$$

since  $\alpha \rightarrow \int_t^T d_t(s) U_{t,\alpha}(s) ds$  is surjective. Thus, by (8) and (9)

$$|v(x, t) - d_t(T)g(x)| \leq Cd_t(T)\delta_t$$

for an appropriated constant  $C$ . Given that  $d_t(T)$  tends to 1 and  $\delta_t$  tends to 0 as  $t \rightarrow T$ , we are done.  $\square$

## 4. DYNAMIC PROGRAMMING EQUATION

For any  $y \in AC_{x,t}$  and  $t \leq \tau \leq T$ , let  $Y_\tau$  denote an optimal arc from initial condition  $(y(\tau), \tau)$ , that is,

$$Y_\tau(s) = Y_{y(\tau), \tau, \alpha(y(\tau), \tau)}(s),$$

which exists by Corollary 3.1.

Consider for  $t < \tau \leq T$  the function

$$W(x, t, \tau) = \int_\tau^T (d_t(s) - d_\tau(s)) \ell(\dot{Y}_\tau(s)) ds + (d_t(T) - d_\tau(T)) g(Y_\tau(T))$$

**Lemma 4.1.** *For every initial condition  $(x, t)$ , admissible arc  $y \in AC_{x,t}$  and  $t \leq \tau \leq T$ , we have*

$$(10) \quad v(x, t) \leq \int_t^\tau d_t(s) \ell(\dot{y}(s)) ds + v(y(\tau), \tau) + W(x, t, \tau).$$

*Proof.* Let  $y \in AC_{x,t}$  be fixed but arbitrary. If  $\tau = T$ , then  $y(T) = Y_T(T)$  and  $v(y(T), T) = d_T(T)g(y(T)) = g(Y_T(T))$ . Then (10) reduces to  $v(x, t) \leq \int_t^T d_t(s) \ell(\dot{y}(s)) ds + d_t(T)g(y(T))$ , which is true by the definition of  $v$ . Now, suppose  $\tau < T$ . Let  $\alpha(y(\tau), \tau) \in A(y(\tau), \tau)$ . Then, by Corollary 3.1

$$\int_\tau^T d_\tau(s) \ell(\dot{Y}_\tau(s)) ds + d_\tau(T)g(Y_\tau(T)) = v(y(\tau), \tau).$$

Let us define the admissible arc  $\tilde{y} \in AC_{x,t}$  by

$$\tilde{y}(s) = \begin{cases} y(s), & \text{if } t \leq s \leq \tau; \\ Y_\tau(s), & \text{if } \tau < s \leq T. \end{cases}$$

We have

$$\begin{aligned} v(x, t) &\leq \int_t^T d_t(s) \ell(\dot{\tilde{y}}(s)) ds + d_t(T)g(\tilde{y}(T)) \\ &= \int_t^\tau d_t(s) \ell(\dot{y}(s)) ds + \int_\tau^T d_\tau(s) \ell(\dot{Y}_\tau(s)) ds + d_\tau(T)g(Y_\tau(T)) \\ &\quad + \int_\tau^T (d_t(s) - d_\tau(s)) \ell(\dot{Y}_\tau(s)) ds + (d_t(T) - d_\tau(T))g(Y_\tau(T)) \\ &= \int_t^\tau d_t(s) \ell(\dot{y}(s)) ds + v(y(\tau), \tau) \\ &\quad + \int_\tau^T (d_t(s) - d_\tau(s)) \ell(\dot{Y}_\tau(s)) ds + (d_t(T) - d_\tau(T))g(Y_\tau(T)). \end{aligned}$$

□



**Corollary 4.1.** (Dynamic Programming). *For every initial  $(x, t)$  and  $t \leq \tau \leq T$*

$$(11) \quad v(x, t) = \min_{y \in \text{AC}_{x,t}} \left\{ \int_t^\tau d_t(s) \ell(\dot{y}(s)) ds + v(y(\tau), \tau) \right\} + W(x, t, \tau).$$

*Proof.* In fact (10) is an equality since an optimal arc is attained for every initial condition  $x, t$ , by Corollary 3.1.  $\square$

Now consider the function

$$(12) \quad w(x, t, \alpha) = - \int_t^T \frac{\partial d_t}{\partial t}(s) \ell(U_{t,\alpha}(s)) ds - \frac{\partial d_t}{\partial t}(T) g(Y_{x,t,\alpha}(T)).$$

The (generalized) dynamic programming equation is as follows. It could be obtained for a more general optimal control problem with some additional assumptions.

**Theorem 4.1.** (Dynamic Programming Equation). *Suppose that for every  $t \leq s \leq T$ ,  $d_t(s)$ ,  $(\partial/\partial t)d_t(s)$  are continuous in  $t$  and summable in  $s$ . Let  $(x, t)$  be a point at which the value function  $v$  is differentiable. Then:*

$$(13) \quad -v_t(x, t) + \ell^*(-v_x(x, t)) + w(x, t, \alpha(x, t)) = 0, \text{ in } \mathbb{R}^n \times (0, T).$$

*Proof.* 1. By Lemma 4.1, if  $t + h < T$

$$(14) \quad \begin{aligned} v(x, t) - v(y(t+h), t+h) &\leq \int_t^{t+h} d_t(s) \ell(\dot{y}(s)) ds \\ &+ \int_{t+h}^T (d_t(s) - d_{t+h}(s)) \ell(\dot{Y}_{t+h}(s)) ds \\ &+ (d_t(T) - d_{t+h}(T)) g(Y_{t+h}(T)), \end{aligned}$$

for any  $y \in \text{AC}_{x,t}$ .

2. The correspondence  $A$  is compact valued and upper semicontinuous, hence we can assume  $\lim_{h \rightarrow 0^+} \alpha(y(t+h), t+h) \in A(x, t)$ ; we denote the limit by  $\alpha(x, t)$ . By continuity  $\lim_{h \rightarrow 0^+} Y_{t+h}(s) = Y_t(s)$ , and  $\lim_{h \rightarrow 0^+} \dot{Y}_{t+h}(s) = \dot{Y}_t(s)$ . Then,

$$\lim_{h \rightarrow 0^+} h^{-1} (d_t(T) - d_{t+h}(T)) g(Y_{t+h}(T)) = - \frac{\partial d_t}{\partial t}(T) g(Y_t(T)),$$

and

$$\lim_{h \rightarrow 0^+} h^{-1} \int_{t+h}^T (d_t(s) - d_{t+h}(s)) \ell(\dot{Y}_{t+h}(s)) ds = - \int_t^T \frac{\partial d_t}{\partial t}(s) \ell(\dot{Y}_t(s)) ds.$$

3. Taking limits in (14)

$$\begin{aligned} \lim_{h \rightarrow 0^+} h^{-1} (v(x, t) - v(y(t+h), t+h)) &\leq \lim_{h \rightarrow 0^+} h^{-1} \int_t^{t+h} d_t(s) \ell(\dot{y}(s)) ds \\ &- \lim_{h \rightarrow 0} h^{-1} W(x, t, t+h) \end{aligned}$$

for every  $y \in \text{AC}_{x,t}$ . This yields

$$-v_t(x, t) - v_x(x, t) \cdot u - \ell(u) + w(x, t, \alpha(x, t)) \leq 0$$

for every  $u \in \mathbb{R}^n$ . Recalling the definition of  $\ell^*$ , this is equivalent to

$$-v_t(x, t) + \ell^*(-v_x(x, t)) + w(x, t, \alpha(x, t)) \leq 0.$$

4. To prove the equality, we use the same argument. Notice that equality holds in (14) for  $y(s) = Y_t(s)$ .  $\square$

**Remark 4.1.** *If  $\frac{\partial d_t}{\partial t}(s) = \rho(t)d_t(s)$  for some continuous function  $\rho$ , then equation (12) gives  $w(x, t, \alpha(x, t)) = \rho(t)v(x, t)$  hence, (13) takes the form of a Hamilton–Jacobi equation with a dissipation term*

$$-v_t(x, t) + \ell^*(-v_x(x, t)) + \rho(t)v(x, t) = 0, \text{ in } \mathbb{R}^n \times (0, T).$$

*This happens if and only if (3) holds, since we are assuming  $d_t(t) = 1$  for each  $t$ .*

In the general case, the dynamic programming equation (13) has a complicated structure. Indeed, the optimal arc itself enters the formulation as a non–local term, thus the applicability of the equation should be taken with caution. In contrast, the solution given in (4) is simpler. This stresses the usefulness of having a Hopf–Lax formula at hand. Nevertheless, we can give a more amenable form to the dynamic programming equation, close to classical standards, when assuming that both the value function and function  $g$  are differentiable. This is the content of the next theorem.

**Theorem 4.2.** *With the same assumptions as in Theorem 4.1, assume further that  $\nabla^2 \ell(u)$  is definite positive for every  $u \in \mathbb{R}^n$  and that  $g$  is differentiable; then, the dynamic programming equation (13) is*

$$(15) \quad -v_t(x, t) + \ell^*(-v_x(x, t)) + w(x, t, \iota(-v_x(x, t))) = 0, \text{ in } \mathbb{R}^n \times (0, T).$$

*Proof.* Since we are supposing  $v$  is differentiable, Danskin’s Theorem applied to (4) gives

$$v_x(x, t) = d_t(T) \nabla g(Y_{x,t,\alpha(x,t)}(T)).$$

On the other hand,  $\alpha$  is an unrestricted minimum of  $V$ , hence

$$\begin{aligned} 0 &= V_\alpha(x, t, \alpha) \\ &= \left( \nabla \ell(\alpha) + d_t(T) \nabla g(Y_{x,t,\alpha}(T)) \right) \left( \int_t^T d_t^{-1}(s) \nabla \iota(d_t^{-1}(s) \nabla \ell(\alpha)) ds \right) \nabla^2 \ell(\alpha) \\ &= \left( \nabla \ell(\alpha) + v_x(x, t) \right) \left( \int_t^T d_t^{-1}(s) \nabla \iota(d_t^{-1}(s) \nabla \ell(\alpha)) ds \right) \nabla^2 \ell(\alpha). \end{aligned}$$

Since  $\nabla^2 \ell$  has maximal rank and  $\nabla \iota(\cdot) = (\nabla^2 \ell(\cdot))^{-1}$ , the gradient of  $V$  with respect to  $\alpha$  is the null vector only if  $\nabla \ell(\alpha) = -v_x(x, t)$  and then  $\alpha(x, t) = \iota(-v_x(x, t))$  at points of differentiability of  $v$  (incidentally, this shows that  $\alpha$  must be unique at points of differentiability of  $v$ ). Plugging this value for  $\alpha$  into  $w(x, t, \alpha)$  we reach the expression for the dynamic programming equation asserted in the theorem.  $\square$

## 5. REGULARITY OF THE VALUE FUNCTION

By Rademacher's Theorem, a locally Lipschitz function is almost everywhere differentiable. Thus, by Theorem 3.2, the value function  $v$ , which is characterized by (4) also satisfies the dynamic programming equation almost everywhere. Summarizing:

**Theorem 5.1.** *With the same assumptions as in Theorem 4.2, the function  $v$  defined by the Hopf-Lax formula (4) is the value function (2), which is locally Lipschitz continuous in  $\mathbb{R}^n \times [0, T)$ , and solves the terminal value problem (in a generalized sense)*

$$(16) \quad \begin{cases} -v_t + \ell^*(-v_x) + w(x, t, \iota(-v_x)) = 0, & a.e. \in \mathbb{R}^n \times (0, T); \\ v = g, & \text{on } \mathbb{R}^n \times \{t = T\}. \end{cases}$$

In the conditions of the above theorem, for the particular case of the Hamilton-Jacobi equation with dissipation we have

**Corollary 5.1.** *Function  $v$  given by (4) with  $d_t(s) = \exp(-\int_t^s \rho(r) dr)$ ,  $t \leq s \leq T$  is locally Lipschitz continuous in  $\mathbb{R}^n \times [0, T)$ , and solves the terminal value problem (1).*

Now we establish some results on the smoothness of the value function.

**Theorem 5.2.** *With the same assumptions as in Theorem 4.2, suppose further that  $g$  is convex.*

- (1) *If  $g$  is of class  $C^1$ , then the value function  $v$  is differentiable in  $\mathbb{R}^n \times (0, T)$  and the minimizer  $\alpha$  is continuous.*
- (2) *If  $g$  is of class  $C^2$  and  $\nabla g$  is bounded, then the value function  $v$  is also of class  $C^2$  in  $\mathbb{R}^n \times (0, T)$  and the minimizer  $\alpha$  is of class  $C^1$  in  $\mathbb{R}^n \times (0, T)$ .*

*Proof.* The minimizers  $\alpha$  in (4) satisfy

$$(17) \quad \nabla \ell(\alpha) + d_t(T) \nabla g(Y_{x,t,\alpha}(T)) = 0,$$

as shown in the proof of Theorem 4.2.

1. For  $x, t$  fixed but arbitrary, the mapping  $\alpha \mapsto \nabla \ell(\alpha) + d_t(T) \nabla g(Y_{x,t,\alpha}(T))$  is strictly monotone due to the convexity of  $g$  and the strict convexity of  $\ell$ , thus  $\alpha(x, t)$  is unique; as a correspondence,  $\{\alpha(x, t)\}$  is upper semicontinuous thus, as a function it is continuous. The uniqueness of  $\alpha$  leads to the differentiability of the value function, by Danskin's Theorem.

2. The derivative of the L.H.S. of (17) with respect to  $\alpha$  is

$$\left( I + d_t(T) \nabla^2 g(Y_{x,t,\alpha}(T)) \int_t^T d_t^{-1}(s) \nabla \iota(d_t^{-1}(s) \nabla \ell(\alpha)) ds \right) \nabla^2 \ell(\alpha),$$

with  $I$  being the identity matrix. Given our assumptions, this vector has norm  $\geq 1$  hence, (17) locally defines  $\alpha(x, t)$  of class  $C^1$ . This function is defined globally since the mapping  $\alpha \mapsto \nabla \ell(\alpha) + d_t(T) \nabla g(Y_{x,t,\alpha}(T))$  is proper because

$$\lim_{\alpha \rightarrow \pm\infty} \left( \nabla \ell(\alpha) + d_t(T) \nabla g(Y_{x,t,\alpha}(T)) \right) = \pm\infty.$$

By Danskin's Theorem,  $v_x(x, t) = d_t(T) \nabla g(Y_{x,t,\alpha(x,t)}(T))$  is of class  $C^1$  hence, by the dynamic programming equation (15),  $v_t$  is also of class  $C^1$ .  $\square$

Finally, a result concerning the monotonic behavior of  $\alpha$  in the scalar case.

**Theorem 5.3.** *With the same assumptions as in Theorem 4.1 with  $n = 1$ , assume further that  $g$  is convex and of class  $C^2$ .*

- (1) *For each time  $0 < t < T$ , there exists for all but at most countably many values of  $x \in \mathbb{R}$  a unique point  $\alpha(x, t)$  where the minimum in (4) is attained.*
- (2) *The mapping  $x \mapsto \alpha(x, t)$  is nondecreasing.*

*Proof.* Since  $v$  is locally Lipschitz, it is differentiable almost everywhere, thus the minimum  $\alpha$  is unique almost everywhere. On the other hand, the crossed derivative of  $V(x, t, \alpha)$  with respect to  $x$  and  $\alpha$  is

$$d_t(T) \nabla^2 \ell(\alpha) \nabla^2 g(Y_{x,t,\alpha}) \int_t^T d_t^{-1}(s) \nabla \iota(d_t^{-1}(s) \nabla \ell(\alpha)) ds \geq 0.$$

Hence,  $(x, \alpha) \mapsto V(x, t, \alpha)$  is supermodular, and by Topkis' Theorem, [8],  $A(x, t)$  is a nonempty compact sublattice which admits a lowest element, which we denote again by  $\alpha(x, t)$ , satisfying  $\alpha(x_2, t) \geq \alpha(x_1, t)$  whenever  $x_2 > x_1$ . Then the mapping  $x \mapsto \alpha(x, t)$  is non-decreasing and thus continuous for all but at most countably many  $x$ .  $\square$

## REFERENCES

- [1] Barro, R., Ramsey meets Laibson in the neoclassical growth model, *Quat. J. Econ.* **114** (1999), 1125–1152.
- [2] Barron, E.N., R. Jensen, W. Liu, Hopf–Lax–type formula for  $u_t + H(u, Du) = 0$ , *J. Diff. Equat.* **126** (1996), 48–61.
- [3] Evans, L.C., *Partial differential equations*, American Mathematical Society, Rhode Island, 1998.
- [4] Karp L., Non-constant discounting in continuous time, *J. Econ. Theory* **132** (2007), 557–568.
- [5] Hopf, E., Generalized solutions of nonlinear equations of first order, *J. Math. Mech.* **14** (1965), 951–973.
- [6] Lax, P.D., Hyperbolic systems of conservation laws II, *Commun. Pure Appl. Math.* **10** (1957), 537–566.

- [7] Marín-Solano, J., J. Navas, Non-constant discounting in finite horizon: The free terminal time case, *J. Econ. Dyn. Control* **33** (2009), 666-675.
- [8] Topkis, D., Minimizing a submodular function on a lattice, *Operations Research* **26** (1978), 305-321.
- [9] Van, T.D., M. Tsuji, N.D. Thai Son, *The characteristic method and its generalizations for first-order nonlinear partial differential equations*, Chapman & Hall, London, 2000.

Juan Pablo Rincón-Zapatero  
Departamento de Economía  
Universidad Carlos III de Madrid,  
E-28903 Getafe, Spain  
*E-mail:* jrincon@eco.uc3m.es