A Combined Model of Classical and Impulse Controls for Emission and Stock Abatement Policies*

Akira Maeda†
College of Arts and Sciences, The University of Tokyo
Motoh Tsujimura‡
Faculty of Commerce, Doshisha University

January 28, 2013

Abstract

This paper investigates emission and stock abatement policy decisions by the use of a stochastic optimal control model with a continuous-time setting. Environmental policies are typically referred as emission flow abatement (reduction) activities and thus, can be formulated as classical optimal control problems. However, because of recent shifts in the focus of policy measures, it is getting recognized that direct control of stock variables is also possible. To formulate such direct stock control in the framework of stochastic optimal control theory, an advanced form of control, called impulse control, is necessary. The purpose of the paper is to develop a model that combines classical flow control and impulse control, and to examine its mathematical features to obtain some policy implications.

Keywords: emission and stock control: stochastic control

1 Introduction

Major environmental challenges including climate change, acid rain, etc., can be ultimately viewed as problems of how to reduce negative effects of accumulation of environmental substances (e.g. greenhouse gases: GHGs, or pollutants) under the circumstance that there are many constraints along time horizon. In mathematical terms, they are formulated as dynamic optimization problems: Flow and stock variables are defined to describe dynamical features of accumulation of environmental substances; negative effects are represented by objective functions to be minimized. When the problem is formulated in a continuous-time setting with some stochastic processes, it is called stochastic optimal control. In a standard setting of such stochastic optimal control, the control is assumed to be exercised on flow variables. In the case of climate change, in fact, it is typically assumed that by controlling emissions, one can change

*This research was supported in part by the Environment Research and Technology Development Fund (S10-1(3)) of the Ministry of the Environment, Japan and the Grant-in-Aid for Scientific Research (No. 24510213) of the Japan Society for the Promotion of Science.

†3-8-1 Komaba, Meguro, Tokyo 153-8902, Japan; +81 3 5465 7740; maeda@global.c.u-tokyo.ac.jp

‡Kamigyo-ku, Kyoto, 602-8580 Japan; +81 75 251 4582; mtsujimu@mail.doshisha.ac.jp

the accumulation of GHGs to minimize (maximize) the net present value of sum of social costs (benefits, respectively).

While the assumption of control by flow variables (i.e., GHG emissions) has long been considered to be realistic in the climate change policy debate, it is being relaxed due to recent shifts in the focus of the policy debate. An example of shifts in the focus is that "adaptation" is drawing more attentions than ever. Some measures of adaptation include building higher seawalls against the rise of the sea level, reallocating pieces of farmland to cope with the changes in meteorological conditions, etc. Because the sea level and meteorological elements are related directly to GHG concentration in the atmosphere, these adaptation measures can be viewed as direct control of stock variables.

Another example of shifts in the focus is a growing attention to geoengineering or climate engineering. For instance, some advocates propose engineering methods of direct reduction of carbon dioxide in the atmosphere (some methods are called carbon dioxide removal, CDR). Others propose projects called solar radiation management (SRM) including the creation of stratospheric sulfate aerosols. These methods are intended to directly manipulate GHG concentration in the atmosphere and/or its effect, and thus can be viewed as another form of direct control of stock variables.

Notice that a form of control by stock variables is completely different from that of control by flow variables in that the former usually requires large amount of expense (or cost) at one time for each control. Because of this feature, control by stock variables is represented by discontinuous and countable shifts of these variables, which is called "impulse" in optimal control theory. In short, recent shifts in the focus of the climate policy debate are calling for some mathematical treatment of impulses on stock variables.

The purpose of this paper is to introduce an advanced form of stochastic optimal control into the mathematical formulation of environmental policy with extended measures. The advanced form addresses the combination (or mix) of impulse control with classical flow control. More specifically, we develop a model that combines these two control forms in the context of environmental policy and examine its analytical features to obtain some policy implications.

The paper is organized as follows: Section 2 addresses the model. Section 3 discusses necessary and sufficient conditions for the optimality. Section 4 investigates the solution. Section 5 presents the results of numerical analysis. Section 6 concludes our discussion.

2 The Model

Let η_t and Y_t denote carbon (or pollutants, in general) emission at time t and carbon stock in the atmosphere at time t, respectively. We assume that the dynamics is described by the following continuous-time stochastic differential equation:

$$\mathrm{d}Y_t = (\gamma\eta_t - \delta Y_t)\mathrm{d}t + \sigma Y_t \mathrm{d}W_t, \qquad Y_0 = y \in \mathbb{R}_+, \tag{2.1}$$

where $\gamma \in (0,1)$ is the emission coefficient, $\delta \in (0,1)$ is the constant depreciation rate of the stock and $\sigma > 0$ is its volatility. W_t is a standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$, where $\mathcal{F}_t$ is generated by W_t in $\mathbb{R}$, i.e., $\mathcal{F}_t = \sigma(W_s, s \leq t)$. The agent can decide how much he/she discharges, that is, emission η_t is a control variable. We assume that emission process is a feedback control in the sense that $\eta_t = \eta(t, Y_t)$ and $\eta(t, Y_t)$ is continuous.

The economy is assumed to suffer from the effects due to the carbon stock. The damage function $D(Y_t)$ is assumed to be strictly convex and be specified as a form of:

$$D(Y_t) = aY_t^b, \tag{2.2}$$

where $a > 0$ is the damage conversion factor and $b > 1$ is a damage elasticity of stock. Reducing carbon emission flow requires emission abatement cost. Let $C(\eta_t)$ denote the abatement cost function and be given by:

$$C(\eta_t) = c(\overline{\eta} - \eta_t)^d, \tag{2.3}$$

where $c > 0$ is the emission flow abatement conversion factor, $d = \{2, 4, 6, \cdots\}$ is the emission flow abatement cost parameter and $\overline{\eta}$ is the emission level before the agent reduces the emission[1]. In addition, we assume that it is possible to directly control the carbon stock with a fixed cost at every time of the control. More specifically, we introduce the following cost function for the direct control of the stock variable Y_t:

$$K(\zeta) = k_0 + k_1\zeta, \tag{2.4}$$

where $k_0 > 0$ is the fixed cost, $k_1 > 0$ is the proportional cost coefficient, and $\zeta_i := Y_{\tau_i -} - Y_{\tau_i}$ with $\tau_i \leq t < \tau_{i+1} < \infty$, $i \geq 0$. τ_i stands for the ith time of discretized stock controls. Then, the stock abatement policy is defined by a pair of the size of abatement and its timing:

$$v = \{(\tau_i, \zeta_i)\}_{i \geq 0}. \tag{2.5}$$

That is, the stock abatement policy is defined as an impulse control.

The agent's total cost function J is expressed as follows:

$$J(y; w) = \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-rt}[D(Y_t) + C(\eta_t)]\mathrm{d}t + \sum_{i=0}^{\infty} \mathrm{e}^{-r\tau_i} K(\zeta_i)\mathbf{1}_{\{\tau_i < \infty\}}\right]. \tag{2.6}$$

where r is a discount rate and w is a combined emission and stock abatement policy which is defined by a combined classical and impulse control:

$$w = (\eta, v) = (\{\eta_t\}_{t \geq 0}, \{(\tau_i, \zeta_i)\}_{i \geq 0}). \tag{2.7}$$

Hereafter to simplify the expression, we use combined abatement policy not combined emission and stock abatement policy.

We define the set of admissible combined abatement policies as follows:

Definition 2.1 (Admissible Combined Abatement Policies). *A combined abatement policy w is admissible if the followings are satisfied:*

$$\mathbb{E}\left[\int_0^\infty \mathrm{e}^{-rt} D(Y_t)\mathrm{d}t\right] < \infty; \tag{2.8}$$

$$0 \leq \tau_i \leq \tau_{i+1}, \quad a.s. \quad i \geq 0; \tag{2.9}$$

$$\tau_i \text{ is an } \{\mathcal{F}_t\}_{t \geq 0}\text{-stopping time}, \quad i \geq 0; \tag{2.10}$$

$$\zeta_i \text{ is } \mathcal{F}_{\tau_i}\text{-measurable}, \quad i \geq 0; \tag{2.11}$$

$$\Pr\left[\lim_{i \to \infty} \tau_i \leq \hat{T}\right] = 0, \quad \hat{T} \in [0, \infty). \tag{2.12}$$

[1]Baudry (2000) investigates emission flow abatement policy. He formulated quadratic-type damage function and emission flow abatement function.

The condition given by eq. (2.12) means that stock abatement policy will only occur finitely before a terminal time, $\hat{T}$. See, for example, Cadenillas and Zapatero (2000). Let $\mathcal{W}$ be the set of admissible combined abatement policies.

The agent problem is to choose the combined abatement policy w in order to minimize the expected total cost J and given by:

$$V(y) = \inf_{w \in \mathcal{W}} J(y; w) = J(y; w^*), \tag{2.13}$$

where V is the value function of the agent problem and w^* is an optimal combined abatement policy.

3 Quasi-Variational Inequalities

In the previous section, we formulated the agent's problem as a combined classical and impulse controls problem. From that formulation, we naturally guess that, under an optimal combined abatement policy, the agent reduces emissions at each time and reduces stock of pollutant whenever the stock of pollutant process $Y = \{Y_t\}_{t \geq 0}$ reaches a threshold. In order to verify this conjecture, we prove that a policy induced by the quasi-variational inequalities is optimal combined abatement policy for the agent's problem (2.13).

Suppose that $\phi : \mathbb{R}_{++} \to \mathbb{R}_{++}$ is a continuous function. Let $\mathcal{M}$ denote the stock abatement operator on the space of functions ϕ defined by:

$$\mathcal{M}\phi(y) = \inf_{\zeta \in (0,y)} \{\phi(y - \zeta)) + K(\zeta)\}. \tag{3.1}$$

We assume that ϕ is a twice continuously differentiable function on $\mathbb{R}$, C^2, except on the boundary of the considered region. Let us define an operator $\mathcal{L}^\eta$ of the Y by

$$\mathcal{L}^\eta \phi(y) = \frac{1}{2}\sigma^2 y^2 \phi''(y) + (\gamma\eta - \delta y)\phi'(y) - r\phi(y). \tag{3.2}$$

Since ϕ is not necessarily C^2 in the whole region, we introduce the concept called stochastically C^2. The Green measure of Y^w, $\mathbb{G}(\cdot, y; w)$, defined by

$$\mathbb{G}(\Xi, y; w) = \mathbb{E}\left[\int_0^\infty Y_t^w 1_\Xi \mathrm{d}t\right], \tag{3.3}$$

where 1_Ξ is the indicator of a Borel set $\Xi \subset \mathbb{R}$. ϕ is called *stochastically* C^2 with respect to Y^w if $\mathcal{L}^\eta\phi(y)$ is well defined point wise for almost all y with respect to the Green measure $\mathbb{G}(\cdot, y; w)$. See Brekke and Øksendal (1998) for more details of the Green measure. Henceforth, we assume that ϕ is stochastically C^2 with respect to Y^w.

We are now in a position to define the quasi-variational inequalities (QVI).

Definition 3.1 (QVI)**.** *The following relations are called the QVI for the agent's problem (2.13):*

$$\mathcal{L}^\eta \phi(y) + D(y) + C(\eta) \geq 0; \tag{3.4}$$

$$\phi(y) \leq \mathcal{M}\phi(y); \tag{3.5}$$

$$\left[\min_\eta [\mathcal{L}^\eta \phi(y) + D(y) + C(\eta)]\right] [\mathcal{M}\phi(y) - \phi(y)] = 0. \tag{3.6}$$

Equation (3.6) is the complementary condition and is able to be rewritten as

$$\min_{\eta}[\mathcal{L}^{\eta}\phi(y) + D(y) + C(\eta)] = 0, \quad y \in \mathcal{C} \tag{3.7}$$

and

$$\phi(y) - \mathcal{M}\phi(y) = 0, \quad y \in \mathcal{I}, \tag{3.8}$$

where $\mathcal{C}$ is the continuation region defined by

$$\mathcal{C} := \left\{ y; \phi(y) < \mathcal{M}\phi(y) \text{ and } \min_{\eta}[\mathcal{L}^{\eta}\phi(y) + D(y) + C(\eta)] = 0 \right\} \tag{3.9}$$

and $\mathcal{I}$ is the stock abatement region defined by

$$\mathcal{I} := \left\{ y; \phi(y) = \mathcal{M}\phi(y) \text{ and } \min_{\eta}[\mathcal{L}^{\eta}\phi(y) + D(y) + C(\eta)] > 0 \right\}. \tag{3.10}$$

We define the policy which is derived from the QVI.

Definition 3.2 (QVI policy)**.** *Let ϕ be a solution of the QVI. Then, the following combined abatement policy $\hat{w}$ is called a QVI policy:*

$$(\hat{\eta}_0, \hat{\tau}_0, \hat{\zeta}_0) = (\hat{\eta}_0, 0, 0); \tag{3.11}$$

$$\hat{\tau}_i = \inf\{t \geq \hat{\tau}_{i-1}; Y_t^{\hat{w}} \notin \mathcal{C}\}; \tag{3.12}$$

$$\hat{\zeta}_i = \arg\min \left\{ \phi\left(Y_{\hat{\tau}_i^-}^{\hat{w}} - \zeta_i\right) + K(\zeta_i) : \zeta_i \right\}. \tag{3.13}$$

In this context, $\hat{w}$ is defined by $\hat{w} = (\{\hat{\eta}_t\}_{t\geq 0}, \{(\hat{\tau}_i, \hat{\zeta}_i)\}_{i\geq 0})$ and $Y_t^{\hat{w}}$ is the result of applying the combined abatement policy $\hat{w}$.

We can now prove that a QVI policy is an optimal combined abatement policy. The following theorem is well-known verification theorem. See, for example, Brekke and Øksendal (1998) and Cadenillas and Zapatero (2000)[2].

Theorem 3.1. *(I) Let ϕ be a solution of the QVI and satisfy the following:*

$$\phi \text{ is stochastically } C^2 \text{ w.r.t. } Y^w; \tag{3.14}$$

$$\lim_{t\to\infty} \mathrm{e}^{-rt}\phi(Y_t^w) = 0, \quad a.s. \quad w \in \mathcal{W}; \tag{3.15}$$

$$\text{the family } \{\phi(Y_\tau^w)\}_{\tau<\infty} \text{ is uniformly integrable w.r.t. } \mathbb{P} \quad w \in \mathcal{W}. \tag{3.16}$$

Then, we obtain

$$\phi(y) \leq J(y; w) \quad w \in \mathcal{W}. \tag{3.17}$$

[2]Cadenillas and Zapatero (2000) and Mundaca and Øksendal (1998) investigate control problem of foreign currency exchange by using combined absolute continuous control and impulse control methods.

(II) From (3.4) – (3.6) and (3.9), we have

$$\mathcal{L}^{\hat{\eta}}\phi(y) + D(y) + C(\hat{\eta}) = 0, \quad y \in \mathcal{C}. \tag{3.18}$$

Suppose $\hat{w} \in \mathcal{W}$, i.e., the combined abatement policy is admissible. Then, we obtain

$$\phi(y) = J(y; \hat{w}). \tag{3.19}$$

Hence, we have

$$\phi(y) = V(y) = J(y; \hat{w}). \tag{3.20}$$

Therefore, $\hat{w}$ is optimal.

Proof. See Appendix. □

4 Optimal Combined Abatement Policy

From the formulation of the agent's problem (2.13), it is able to guess that, under an optimal combined abatement policy, the agent reduces the stock of pollutant whenever the stock of pollutant Y^w reaches a level $\overline{y}$, so that it instantaneously decreases to another level $\check{y}$. Hence the agent always reduce the stock of pollutant by $\overline{y} - \check{y}$ at each abatement times, τ_i $(i = 1, 2, \cdots)$.

Let an optimal combined abatement policy be denoted by $w^* = (\{\eta_t^*\}_{t\geq 0}, \{(\tau_i^*, \zeta_i^*)\}_{i\geq 0})$, characterized by $\overline{y}$, $\check{y}$ with $0 < \check{y} < \overline{y}$ such that

$$\tau_i^* := \inf\{t > \tau_{i-1}^*; Y_{t-}^{w^*} \notin \mathcal{C}\}, \tag{4.1}$$

$$\zeta_i^* := Y_{\tau_i^*-}^{w^*} - Y_{\tau_i^*}^{w^*} = \overline{y} - \check{y}, \quad Y_{t-}^{w^*} \notin \mathcal{C}, \tag{4.2}$$

where

$$\mathcal{C} := \{y; y \in (0, \overline{y})\}. \tag{4.3}$$

Therefore, when the agent implements the assumed optimal combined abatement policy w^*, the value function seems to satisfy

$$V(y) = V(\check{y}) + k_0 + k_1(y - \check{y}), \quad y \notin \mathcal{C}. \tag{4.4}$$

Assume that V is differentiable at $y = \overline{y}$. Under w^*, if the initial level of the pollutant y is $y = \overline{y} + \varepsilon$, where $\varepsilon > 0$, then the optimal size of pollutant abatement is $\zeta^* = (\overline{y} + \varepsilon) - \check{y}$. Thus, we have for $y \geq \overline{y}$

$$V(\overline{y} + \varepsilon) = V(\check{y}) + k_0 + k_1[(\overline{y} + \varepsilon) - \check{y}]. \tag{4.5}$$

Substituting y into $\overline{y}$ in (4.4) for $y \geq \overline{y}$ and subtracting from (4.5), we obtain

$$V(\overline{y} + \varepsilon) - V(\overline{y}) = k_1\varepsilon. \tag{4.6}$$

Dividing (4.6) by ε and taking $\lim_{\varepsilon\to 0}$ in (4.6) we obtain

$$V'(\overline{y}) = k_1. \tag{4.7}$$

By (4.1) and (4.2), the agent's expected total cost function, $J(y; w)$, is minimized at $\zeta^* = \overline{y} - \check{y}$. Hence, by the first order condition for the minimization $\mathrm{d}[J(\overline{y} - \zeta) + (k_0 + k_1\zeta)]/\mathrm{d}\zeta\mid_{\zeta=\overline{y}-\check{y}} = 0$, we obtain

$$V'(\check{y}) = k_1. \tag{4.8}$$

Furthermore, we can conjecture that for $Y_t^{w^*} \in \mathcal{C}$

$$\mathcal{L}^{\eta_t^*}\phi(Y_t^{w^*}) + D(Y_t^{w^*}) + C(\eta_t^*) = \min_{\eta}[\mathcal{L}^{\eta}\phi(Y_t^{w^*}) + D(Y_t^{w^*}) + C(\eta)] = 0. \tag{4.9}$$

Then, optimal emission flow is given by

$$\eta_t^* = \arg\min \left\{ \gamma\eta_t\phi'(Y_t^{w^*}) + c(\overline{\eta} - \eta_t)^d \right\}. \tag{4.10}$$

Hence we obtain that

$$\eta_t^* = \overline{\eta} - \left(\frac{\gamma}{cd}\phi'(Y_t^{w^*}) \right)^{\frac{1}{d-1}} \tag{4.11}$$

Substituting (4.11) into (4.9) yields

$$\frac{1}{2}\sigma^2 y^2 \phi''(y) - \gamma \left(\frac{d-1}{d} \right) \left(\frac{\gamma}{cd} \right)^{\frac{1}{d-1}} (\phi'(y))^{\frac{d}{d-1}} + (\gamma\overline{\eta} - \delta y)\phi'(y) - r\phi(y) + ay^b = 0. \tag{4.12}$$

Since this ordinary differential equation is nonlinear, we cannot obtain an explicit solution.

There exist two unknown constant parameters, $\overline{y}$ and $\check{y}$. These parameters are uniquely determined by the following simultaneous equations:

$$\phi(\overline{y}) = \phi(\check{y}) + k_0 + k_1(\overline{y} - \check{y}); \tag{4.13}$$

$$\phi'(\overline{y}) = k_1; \tag{4.14}$$

$$\phi'(\check{y}) = k_1, \tag{4.15}$$

where ϕ is the solution of (4.9) for $y \in (0, \overline{y})$.

5 Numerical Analysis

The results of numerical analysis will be presented in the conference.

6 Conclusion

Because of recent shifts in the focus of policy measures–growing attentions to adaptation activities as well as geoengineering measures–mathematical formulation of environmental policies calls for a more advanced tool. Direct control of stock variables with certain fixed costs can only be formulated as impulse control problems. Being motivated by such recent shift, we developed a model that combines classical flow control and impulse control. We showed that the optimal policy is represented by the set of quasi-variational inequalities (QVIs) and boundaries set for stock variables. We discussed the sufficiency of the optimality conditions by proving a theorem called verification theorem. We also showed a differential equation and a set of simultaneous equations that embodies the optimal policy. Unfortunately, analytical solution for them is not obtainable. Solving these equations numerically is a next step for us to take.

References

Baudry, M. (2000), Joint management of emission abatement and technological innovation for stock externalities, *Environmental and Resource Economics*, **16**, 161–183.

Brekke, K.A. and B. Øksendal, (1998), A Verification Theorem for Combined Stochastic Control and Impulse Control, in: Decreusefond, L. et al. (eds.), *Stochastic Analysis and Related Topics*, Vol.6. Bikhäuser, Basel. 211–220.

Cadenillas, A. and F. Zapatero (2000), Classical and Impulse Stochastic Control of the Exchange Rate Using Interest Rates and Reserves, *Mathematical Finance*, **10**, 141–156.

Mundaca, G. and B. Øksendal, (1998), Optimal Stochastic Intervention Control with Application to the Exchange Rate, *Journal of Mathematical Economics*, **29**, 225–243.

Appendix

Proof. (I) Assume that ϕ satisfies (3.14)-(3.16). Choose $w \in \mathcal{W}$. Let $\theta_{i+1} := \tau_i \vee (\tau_{i+1} \wedge s)$ for any s (≥ 0). Then, by the generalized Dynkin formula, we obtain

$$\mathbb{E}\left[\mathrm{e}^{-r\theta_{i+1}^{-}}\phi(Y^w_{\theta_{i+1}})\right] = \mathbb{E}\left[\mathrm{e}^{-r\tau_i}\phi(Y^w_{\tau_i})\right] + \mathbb{E}\left[\int_{\tau_i}^{\theta_{i+1}^{-}} \mathrm{e}^{-rt}\mathcal{L}\phi(Y^w_t)\mathrm{d}t\right]. \tag{A.1}$$

Hence, from (3.4) we obtain

$$\mathbb{E}\left[\mathrm{e}^{-r\theta_{i+1}^{-}}\phi(Y^w_{\theta_{i+1}})\right] \geq \mathbb{E}\left[\mathrm{e}^{-r\tau_i}\phi(Y^w_{\tau_i})\right] - \mathbb{E}\left[\int_{\tau_i}^{\theta_{i+1}^{-}} \mathrm{e}^{-rt}\pi(Y^w_t)\mathrm{d}t\right], \tag{A.2}$$

where

$$\pi(y^w) = D(y) + C(\eta).$$

Taking $lim_{s\to\infty}$ and using the dominated convergence theorem, we have

$$\mathbb{E}\left[\mathrm{e}^{-r\tau_{i+1}^{-}}\phi(Y^w_{\tau_{i+1}^{-}})\right] \geq \mathbb{E}\left[\mathrm{e}^{-r\tau_i}\phi(Y^w_{\tau_i})\right] - \mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}} \mathrm{e}^{-rt}\pi(Y^w_t)\mathrm{d}t\right]. \tag{A.3}$$

Summing from i=0 to i=m yields

$$\phi(y) + \sum_{i=1}^{m}\mathbb{E}\left[\mathrm{e}^{-r\tau_i}\phi(Y^w_{\tau_i}) - \mathrm{e}^{-r\tau_i^{-}}\phi(Y^w_{\tau_i^{-}})\right] \leq \mathbb{E}\left[\mathrm{e}^{-r\tau_{m+1}^{-}}\phi(Y^w_{\tau_{m+1}^{-}})\right] + \mathbb{E}\left[\int_{\tau_0}^{\tau_{m+1}} \mathrm{e}^{-rt}\pi(Y^w_t)\mathrm{d}t\right]. \tag{A.4}$$

For all $\tau_i < \infty$, following the investment policy, the state process Y^w jumps immediately from $Y^w_{\tau_i^-}$ to a new state level $Y^w_{\tau_i^-} - \zeta_i$. Thus, by (3.1) and $Y^w_{\tau_i^-} - \zeta_i = Y^w_{\tau_i}$, we obtain

$$\phi(Y^w_{\tau_i}) \geq \mathcal{M}\phi(Y^w_{\tau_i^-}) - K(\zeta_i). \tag{A.5}$$

Thus, we have

$$\begin{aligned}\phi(y) \quad &+\sum_{i=1}^{m}\mathbb{E}\left[\left[\mathrm{e}^{-r\tau_i}\mathcal{M}\phi(Y^w_{\tau_i^-})-\mathrm{e}^{-r\tau_i^-}\phi(Y^w_{\tau_i^-})\right]1_{\{\tau_i<\infty\}}\right]\\ \leq \quad &\mathbb{E}\left[\int_0^{\tau_{m+1}}\mathrm{e}^{-rt}\pi(Y^w_t)\mathrm{d}t+\sum_{i=1}^{m}\mathrm{e}^{-r\tau_i}K(\zeta_i)1_{\{\tau_i<\infty\}}+\mathrm{e}^{-r\tau^-_{m+1}}\phi(Y^w_{\tau^-_{m+1}})\right]. \end{aligned} \tag{A.6}$$

It follows from (3.5) that

$$\mathcal{M}\phi(Y^w_{\tau_i^-})-\phi(Y^w_{\tau_i^-})\geq 0. \tag{A.7}$$

Hence, we obtain taht

$$\phi(y)\leq\mathbb{E}\left[\int_0^{\tau_{m+1}}\mathrm{e}^{-rt}\pi(Y^w_t)\mathrm{d}t+\sum_{i=1}^{m}\mathrm{e}^{-r\tau_i}K(\zeta_i)1_{\{\tau_i<\infty\}}+\mathrm{e}^{-r\tau^-_{m+1}}\phi(Y^w_{\tau^-_{m+1}})\right]. \tag{A.8}$$

Taking $lim_{m\to\infty}$ and using (3.15), (3.16) and the dominated convergence theorem, we obtain

$$\phi(y)\leq\mathbb{E}\left[\int_0^{\infty}\mathrm{e}^{-rt}\pi(Y^w_t)\mathrm{d}t+\sum_{i=1}^{m}\mathrm{e}^{-r\tau_i}K(\zeta_i)1_{\{\tau_i<\infty\}}\right]. \tag{A.9}$$

Therefore, (3.17) is proved.

(II) Assume that (3.18) holds and $\hat{w}$ is the QVI policy. Repeat the argument in part (I) for $w=\hat{w}$. Note that inequalities (A.2) through (A.9) become equalities. Thus, we obtain

$$\phi(y)=\mathbb{E}\left[\int_0^{\infty}\mathrm{e}^{-rt}\pi(Y^{\hat{w}}_t)\mathrm{d}t+\sum_{i=1}^{m}\mathrm{e}^{-r\hat{\tau}_i}K(\hat{\zeta}_i)1_{\{\hat{\tau}_i<\infty\}}\right]. \tag{A.10}$$

Hence, we obtain (3.19). Combining (3.19) with (3.17), we obtain

$$\phi(y)\leq\inf_{w\in\mathcal{W}}J(y;w)\leq J(y;\hat{w})=\phi(y). \tag{A.11}$$

Therefore, $\phi(y)=V(y)$ and $w^*=\hat{w}$ is optimal, i.e., the solution of the QVI is the value function and the QVI policy is optimal. The proof is completed.

□