Some problems in Stochastic Control Theory related to Inventory Management and Coarsening

by

Jingchen Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2014

Doctoral Committee:

Professor Xiuli Chao, Co-chair
Professor Joseph Conlon, Co-chair
Professor Erhan Bayraktar
Professor Selim Esedoglu
Professor Edwin Romeijn

To my parents for their support, encouragement and love. Without you, I would not have had the chance to be where I am.

To all my dear friends. Your intelligence inspires me, your loyalty encourages me and your optimism cheers up me.

To my grandpa. Your ever-lasting passion for life and curiosity has always motivated me to expand my horizons.

ACKNOWLEDGEMENTS

In the past five years, I have received help from many people–family members, teachers and friends. Without your help, I would not have finished this dissertation.

I would like to thank Dr. Xiuli Chao for his guidance in bringing me into the research of inventory management. His instruction and encouragement motivated my pursuit in this area. I would like to thank Dr. Joseph Conlon for introducing to me the application of mathematics to another fascinating area. His kind advice has always been more than helpful. I also like to thank my dissertation committee of Dr. Erhan Bayraktar, Dr. Edwin Romeijn and Dr. Sulim Esedoglu for teaching me, guiding me and giving me all kinds of support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

HJB Hamilton-Jacob-Bellman

EOQ Economic Order Quantity

EPQ Economic Production Quantity

PDE Partial Differential Equation

CP Carr-Penrose

LSW Lifshitz-Slyozov-Wagner

ABSTRACT

Some problems in Stochastic Control Theory related to Inventory Management and Coarsening

by

Jingchen Wu

Chair: Xiuli Chao and Joseph Conlon

In this dissertation, we study two stochastic control problems arising from inventory management and coarsening.

First, we study a stochastic production/inventory system with a finite production capacity and random demand. The cumulative production and demand are modeled by a two-dimensional Brownian motion process. There is a setup cost for switching on the production and a convex holding/shortage cost, and our objective is to find the optimal production/inventory control that minimizes the average cost. Both lost-sales and backlogging cases are studied. For the lost-sales model we show that, within a large class of policies, the optimal production strategy is either to produce according to an (s, S) policy, or to never turn on the machine at all (thus it is optimal for the firm to not do the business); while for the backlog model, we prove that the optimal production policy is always of the (s, S) types. Our approach first develops a lower bound for the average cost among a large class of non-anticipating policies, and then shows that the value function of the desired policy reaches the lower bound. The results offer insights on the structure of the optimal control policies as well as the interplay between system parameters.

Then, we study a diffusive Carr-Penrose model which describes the phenomenon of coarsening. We show that the solution and the coarsening rate of the diffusive model converge to the classical Carr-Penrose model. Also, we demonstrate the relationship between the log concavity of the initial condition and the coarsening rate of the system. Under the assumption that the initial condition is log concave, there exists a constant upper bound on the coarsening rate of the diffusive problem. Our approach involves a representation of the solution using Dirichlet Green's function. To estimate this function, we exploit the property of a non-Markovian Gaussian process and derive bounds (both upper and lower) on the ratio between the Dirichlet and the full space Green's functions. The results shed light on the connection between the classical and diffusive Carr-Penrose models, and characterize the coarsening phenomenon under small noise perturbation.

CHAPTER I

Optimal Control of a Brownian Production/Inventory System with Average Cost Criterion

1.1 Introduction and literature review.

A fundamental result in inventory theory is the optimality of (s, S) policy for inventory systems with setup cost (Scarf [36], Veinott [40]). The key assumption for this result is *infinite ordering/production capacity*. That is, regardless of how much is ordered, it will be ready after a leadtime that is independent of the ordering quantity. This assumption is clearly not satisfied in many applications, especially in production systems; all production facilities have finite capacity. Several studies have been conducted attempting to extend the results to the case of finite capacity. In the special case with no setup cost, Federgrun and Zipkin [14] have shown that the optimal strategy for the capacitated inventory system is a simple extension of the optimal base-stock policy to the uncapacitated problem, which is often called the modified base-stock policy. Does such "modification" continue to hold in the case with setup cost? While it is plausible that some form of a modified (s, S) policy would be optimal, it has been shown by several authors, through counterexamples, that this is not true. See for example, Wijngaard [43], and Chen and Lambrecht [5]. Efforts have been made to analyze the structure of the optimal control policy for capacitated inventory system with setup cost, see, e.g., Gallego and Scheller-Wolf [17], and Chen [4], and the best known result is a partial characterization of the optimal control policy. These studies are established for periodic-review production/inventory systems, but similar result holds for continuous-review system when the demand follows a batch Poisson process.

There are also several studies of production/inventory systems using Brownian motion models, and again most of these studies assume infinite production capacity, which means, in the Brownian setting, that the inventory levels can be changed instantaneously. In the stochastic control jargon, this is referred to as *impulse control*. Bather [1] uses Brownian motion to model the demand process and allows the inventory to be controlled instantaneously with setup cost and proportional variable cost; and he shows that an (s, S) policy is optimal under long-run average cost criterion. In [35], Richard considers both infinite and finite horizon problems with discounted cost objective, and he presents sufficient conditions for the optimality of an (d, D, U, u) policy among the class of impulse control policies; this work has been extended in [8] to a more general setting, for which the optimal control is shown to take a form (d, D, U, u). These papers study the backlog inventory model, in which the state variable (the inventory level) takes any real value. Harrison in [21] studies a similar discounted cost optimal control problem of a Brownian model, and he imposes the condition that the state of the system is non-negative. This non-negativity condition leads to a lost-sales inventory control model, and Harrison obtains the optimal impulse control policy for the case with or without setup cost. In [20] Harrison et al. propose another Brownian model for a cash management problem, in which the state of the system can be instantaneously increased or decreased, and the authors show that a $(0, q, Q, S)$ policy is optimal for discounted cost criterion. Sulem [38] investigates the computational issue of the optimal control parameters based on the work of [8]. Ormeci et al. [30] consider linear holding and shortage cost rate and extend the result of [20] to long-run average cost criterion, and they prove that the optimal policy remains the $(0, q, Q, S)$ policy. They also generalize the result to the case when there is finite adjustment condition in the impulse control policies. Dai and Yao [23, 24] further extend the model of Ormeci et al. [30] to convex holding and shortage cost rate, and obtain the optimal impulse controls for both average and discounted costs.

As mentioned, impulse control in Brownian inventory models implicitly assumes infinite production capacity. When the production capacity is finite, which is always true in practice, the inventory levels can only be gradually changed over time at a finite production rate. Normally, the production rate can only be changed among a finite set of alternatives through changing the number of staff, number of shifts, or opening or closing production lines. However, such adjustments can be restricted in some situations. For example, suppose a factory has multiple production lines and, to match the demand rate, the ideal number of lines to run is between n and $n+1$. Then, it is practical to consider only two production alternatives, i.e., n or $n+1$ production lines. We consider another example, in which a factory has only one production line. The factory can decide whether to turn it on or off, instead of changing the production rate. Again the problem is to choose between the two production capacities. The second example is a special case of the first but it captures many practical scenarios. In what follows, we focus on the latter example as the stereotype problem. Therefore, in such situations, the production decisions are when to set up the machine to produce, and when to shut down the production. The resulting optimal production/inventory control problem is regime-switching between production mode and non-production mode. For this optimal switching problem, a fixed setup cost K is incurred whenever the production precess is turned on. Due to this cost, the inventory manager needs to limit the frequency of turning on the production process. Intuitively, the larger the setup cost K, the less frequently the manager should switch to the production mode. In the special case of deterministic demand and production processes, this reduces to the classic EOQ/EPQ model (see e.g., Hax and Candea [22]), in which the inventory manager balances holding/shortage cost with the setup cost to minimize the average cost. This intuition carries over to the stochastic production/demand case. The most plausible stationary control policy is the (s, S) policy: Every time the inventory level reaches or goes above S, the production process is turned off; and as soon as the level drops to or below s, the production process is turned back on; otherwise the production mode remains unchanged. As indicated above, the (s, S) policy has been widely studied and proven optimal among the class of impulse controls for a number of Brownian inventory models.

In this chapter, we study a stochastic production-inventory system with finite production capacity and random demand. The cumulative production when the machine is on, as well as the cumulative demand, are modeled by a two-dimensional Brownian motion process. There is a fixed cost for setting up the machine for production, and there is also convex holding and shortage cost. We are concerned with the optimal production/inventory control strategy that minimizes the long run average cost. Both lost-sales and backlogging cases are studied. For the lost-sales model we show that, within a large class of policies, the optimal production strategy is either to produce according to an (s, S) policy, or never to turn on the machine at all (thus it is optimal for the firm to not enter the business); for the backlog model, we prove that the optimal production policy is always of the (s, S) type. Our approach first develops a lower bound for the average cost among a large class of non-anticipating policies, a powerful method developed in Ormeci et al [30] for an impulse control setting which we generalize here. Then we show that the value function of the proposed policy reaches the lower bound. The results shed lights on the structure of the optimal control policies as well as the interplays between system parameters and their effects on the optimal control parameters and system minimum cost.

The most relevant literature for the present paper is Vickson [42], Doshi [10, 11]. In [42], Vickson considers an average cost production-inventory problem with holding cost rate which is linear of the form $h(x) = h\max\{x, 0\} + p\max\{-x, 0\}$. The production process is deterministic, and the cumulative demand process follows a Brownian motion process. He proves the optimality of the (s, S) policy under certain conditions. In [11], a quadratic cost and discounted cost criterion are assumed. Due to the complexity of the mathematical expression for the cost function, Doshi only proves the optimality of the (s, S) policy for some symmetric cost cases. Both of these two papers study the backlogging model. In [10], Doshi considers a one-dimensional Brownian model with multiple modes and adopts an average cost criterion. He establishes the existence of an optimal stationary policy under the assumption that the inventory level lies in a compact interval with reflecting boundaries, and then he proves the optimality of the (s, S) policy for a symmetric case with quadratic cost function. Puterman [34] considers a Brownian motion model of a storage system, and analyzes the average cost operating under an (s, S) policy; he also investigates the computation of the optimal parameters s and S when the holding cost rate is linear or quadratic. Sheng [37] studies a control problem with discounted cost objective similar to [10] in the sense that it allows the one-dimensional Brownian model to have multiple modes. Sheng provides a sufficient and necessary optimality condition in terms of the value function and shows the optimality of band policies for some special cases.

The problem studied here falls into the category of optimal switching. There have been some recent developments in the understanding of optimal switching problems. Duckworth and Zervos [12] use a dynamic programming principle to derive the Hamilton-Jacobi-Bellman (HJB) equation and apply a verification approach to solve an optimal two-regime switching problem with discounted cost. In [31], Pham uses a viscosity solutions approach to prove the smooth-fit $\mathcal{C}^1$ property of the value functions, and extends the well known results of optimal stopping to one-dimensional optimal switching problem. In [39], Vath and Pham study a two-regime optimal switching problem and provide a partial characterization on the structure of the switching regions. Under the geometric Brownian motion evolution assumption, they explicitly solve the problem for two special cases: 1) identical power profit functions with different diffusion operators; 2) identical diffusion operators with different profit functions. In [32], Pham et al. extend the results for case 2) in [39] to multiple regimes. More recently, Bayraktar and Egami [2] use a sequential approximation method to study a two-regime switching problem with discounted cost criterion, and establish a dynamic programming principle for the value function as two coupled optimal stopping problems. They utilize results from optimal stopping of one-dimensional diffusion processes, and obtain sufficient conditions on the connectedness of the optimal switching region under some specific assumptions. They also obtain simple control policies for several examples where the process under consideration is independent of the control (switching) decision, i.e., the evolution of the process is the same in different switched regions. Ghosh et al., in [18], consider a problem with random switching of modes, in which the dynamic is influenced by the control. They prove the existence of a homogeneous Markov nonrandomized optimal policy using a convex analytic method and the uniqueness of the solution to the HJB equations within a certain class. All the studies above on optimal switching adopt discounted cost criterion, and their approach and results do not directly apply to the average cost case, which is the focus in this chapter.

The optimal production/inventory control problem we study is similar to but more general than that in [42]. We consider a continuous-review production/inventory system, and model the inventory level X_t by a two-dimensional Brownian model process. The necessity of the two-dimensional Brownian motion process stems from the fact that there are uncertainties in both production and demand processes, which cannot be captured with a single-dimension diffusion. The system has two possible modes, for which the diffusion process has different drift and volatility parameters: $(-\mu_0, \sigma_0^2)$ and (μ_1, σ_1^2), with $\mu_0, \mu_1 > 0$. Thus it implies that the production and demand processes do not have to be independent, and are in general correlated. In production mode 0, the machine is idle, so X_t decreases as demand arrives; while in mode 1, production is on and X_t, which represents the difference between the production and demand processes, increases at a net rate μ_1. At any point in time, the inventory manager can switch the mode of the production, which incurs a fixed machine setup cost $K > 0$ if the mode is changed from 0 to 1. In addition, the inventory level X_t incurs a holding and shortage cost rate $h(X_t)$, i.e., if $X_t \geq 0$ then $h(X_t)$ is the holding cost rate and if $X_t < 0$ it is the shortage cost rate. For lost-sales model $X_t \geq 0$ for all t thus $h(X_t)$ only represents holding cost, but there is also a shortage cost for each unit of demand lost. The objective is to control the production process so as to minimize the long-run average cost of the system. In addition, compared with Doshi's work [10], we do not have a finite and compact state space, which leads to additional complexity of the problem as will be seen in Proposition I.4 and Section 1.4.

We first focus on the lost-sales model. In the case of impulse control, $(0, S)$ policy is known to be optimal because of the infinite production capacity. For the finite production capacity model, the manager would possibly start the production before the inventory level hits zero, so as to avoid the possible cost of losing customers. We derive the lower bound for the average cost within a large class of policies, and show that in certain range of the system parameters, there exists a unique optimal (s, S) policy that achieves this lower bound. Thus it establishes the optimality of (s, S) policy. When the system parameters do not fall into that region, we prove that the "never-turn-on-the-machine" is the optimal policy, again within a large class of policies, implying that it is optimal for the firm to not enter the business (or to go out of business). For the backlog models, we show that an (s, S) policy is always optimal within a large class of policies.

There is a technical issue in the verification theorem for optimality of the finite capacity inventory/production problem, which constitutes the major difference between our approach and those in the impulse control papers. In [30], due to the nature of infinite production capacity, the relative value function $f(x)$ for the optimal band policy is guaranteed to be *Lipschitz continuous*, where x is inventory level. As a result, for any control policy with a divergent state X_T, it can be shown that either $E[f(X_T)]$ diverges slower than a linear function of T, leading to a policy inferior to the desired one, or it diverges at least as fast as a linear function of T, incurring an *infinite* average cost. (See the proof of Proposition 2 in [30].) However, this approach fails to work for the finite production capacity model. We will present an example in Section 1.4, which shows the existence of a policy with $E[f(X_T)]$ diverging faster than a linear function of T but yet still incurring *finite* average cost. To overcome this difficulty, in our study we focus on a class of admissible policies, and show that the desired policy is optimal within this class of policies; and we show that this class of policies is large enough to include most policies of practical interest.

There exist abundant papers in the literature studying optimal control of infinite-capacity production/inventory systems, but few on optimal control of finite-capacity production systems, and real world production/inventory systems all have finite capacity. This chapter provides a complete analysis of the optimal control of a capacitated production/inventory system and identifies the optimal control policies. In particular, it offers insights on the range of system parameters under which it is economically optimal for the firm to not enter the business.

The rest of this chapter is organized as follows. In §1.2, the lost-sales model is studied in detail. We present the Brownian motion formulation of the production/inventory problem in §1.2.1. In §1.2.2, we develop a lower bound for the average cost within a large class of policies. In §1.2.3, we identify the optimal parameters s and S within the class of (s, S) policies. Next, in §1.2.4, we show that the system parameters can be divided into two regions. In the first region, the relative value function associated with an (s, S) policy satisfies the lower bound conditions in §1.2.2, thus proving that an (s, S) policy is optimal within a large class of non-anticipating policies. In the other region, we show that the relative value function of the "never-turn-on-machine" policy satisfies all lower bound conditions, proving that the "never-turn-on-machine" policy is optimal. In §1.3, we extend the analysis to the backlogging model, and show that the optimal policy is always an (s, S) policy. We also discuss a special case in which the result can be extended to quasi-convex holding and shortage cost rate function. In §1.4, we discuss the class of policies we have focused on and show that it includes most cases of practical interest.

1.2 The lost-sales model.

In the lost-sales model, any demand arriving during the out of stock period is lost, at a shortage penalty cost. The inventory manager needs to balance the shortage cost, machine setup cost, and the inventory holding cost. We rigorously formulate the problem in the following subsections.

1.2.1 Model and basic assumptions.

We first present the problem formulation. Let Ω be the space of all $\mathbb{R}^2$-valued continuous functions $\omega : [0, \infty) \to \mathbb{R}^2$. Let $B = (B_t^0, B_t^1)_{t \geq 0}$ be a two-dimensional standard Brownian motion under a probability measure P , and $\mathcal{F}_t$ be the natural filtration generated by B. Besides, let $\mathcal{F}$ be a σ-algebra of Ω such that $\mathcal{F}_t \subset \mathcal{F}$ for all $t \geq 0$. Then $(\Omega, \mathcal{F}, P)$ forms the probability space on which the production and demand processes are defined.

Let $W_t^0 = B_t^0$ and $\hat{W}_t = \rho B_t^0 + \sqrt{1-\rho^2} B_t^1$, where $\rho \in [-1, 1]$, then W_t^0 and $\hat{W}_t$ are Brownian motions with correlation coefficient ρ, i.e., $E\left[dW_t^0 \cdot d\hat{W}_t\right] = \text{corr}\left[dW_t^0, d\hat{W}_t\right] = \rho dt$. Denote the demand and production (if it is always on) from time 0 up to time t by D_t and P_t respectively. Suppose D_t and P_t are governed by

$$\begin{aligned} dD_t &= -(-\mu_0 dt + \sigma_0 dW_t^0), \\ dP_t &= \hat{\mu}_1 dt + \hat{\sigma}_1 d\hat{W}_t, \end{aligned}$$

where $\mu_0 > 0$ and $\hat{\mu}_1 > 0$ represent the demand and production rates.

Let X_t denote the inventory level at time t and $Y_t \in \{0, 1\}$ the production mode at time t, which governs the evolution of inventory level. When $Y_t = 1$, the production process is on and the inventory level is affected by both the production and demand processes; and when $Y_t = 0$, the machine is idle, and the inventory level is only affected by the demand process. Due to lost-sales, 0 is a reflecting boundary. Letting Z_t denote the total demand lost up to time t, then the inventory level process is governed by the following stochastic differential equations:

$$\begin{aligned} dX_t &= -\mu_0 dt + \sigma_0 dW_t^0 + dZ_t, \quad \text{if } Y_t = 0; \\ dX_t &= (\hat{\mu}_1 - \mu_0) dt + \sigma_0 dW_t^0 + \hat{\sigma}_1 d\hat{W}_t + dZ_t, \quad \text{if } Y_t = 1. \end{aligned}$$

We note that

$$\begin{aligned} \sigma_0 dW_t^0 + \hat{\sigma}_1 d\hat{W}_t &= (\sigma_0 + \rho\hat{\sigma}_1) dB_t^0 + \sqrt{1-\rho^2}\hat{\sigma}_1 dB_t^1 \\ &= \left(\sigma_0^2 + \hat{\sigma}_1^2 + 2\rho\sigma_0\hat{\sigma}_1\right)^{1/2} dW_t^1 \\ &= \sigma_1 dW_t^1, \end{aligned}$$

where $\sigma_1 := \left(\sigma_0^2 + \hat{\sigma}_1^2 + 2\rho\sigma_0\hat{\sigma}_1\right)^{1/2}$ (here ":=" stands for "defined as"), and W_t^1 is a standard Brownian motion under $(\Omega, \mathcal{F}, P)$ as well. We notice that when $\rho = 0$, the two Brownian motions W_t^0 and $\hat{W}_t$ are independent and $\sigma_1 > \sigma_0$, in that case the variance of the inventory process during production is strictly greater than that during machine idle time; when $\rho < -\hat{\sigma}_1/2\sigma_0$, the variance of the inventory process during the production time is smaller than that during machine idle time. Hereafter, we let $\mu_1 := \hat{\mu}_1 - \mu_0$, so we have

$$dX_t = \mu_1 dt + \sigma_1 dW_t^1 + dZ_t, \quad \text{if } Y_t = 1.$$

The evolution of X_t for different values of Y_t uses W_t^0 and W_t^1 alternatively. It is worth noting that W_t^1 and W_t^0 are two dependent Brownian motion processes.

We have aggregated the production and demand processes during the production mode, which forms a Brownian motion processes with drift $\mu_1 > 0$ and variance parameter σ_1^2; during non-production mode, inventory is only depleted by the demand process, which is a Brownian motion with rate $-\mu_0 < 0$ and variance parameter σ_0^2. Note that the lost sales process Z_t increases only if X_t is equal to 0, and such a process is often referred to as regulated process, see e.g., Harrison [19]. We remark that it is not always optimal for the firm to set the production mode to 1 when X_t hits 0, as will be seen later. We assume that Y_t is right-continuous.

Remark I.1. *In the inventory control literature, it is commonly assumed, in both discrete and continuous time models, that the demand follows a normal distribution (or Brownian motion model for continuous time). See, for example, Section 5.1 of Nahmias [29] for a discrete time model and Section 4 of Gallego [16] for a continuous time model. This is an approximation to reality as there is a positive probability for the demand to take negative values, and indeed, the assumption is made mainly for tractability. The model reflects reality well when the likelihood of generating negative demand is small, e.g., when the average demand is relatively high or the variance of demand is relatively low.*

The state of the system is (X_t, Y_t), with state space $\{(x,y); x \geq 0, y = 0, 1\}$. For an initial state (x, y) of the system and a policy π, we can define a probability measure $P^{\pi}_{x,y}$ and its associated expectation $E^{\pi}_{x,y}$.

Cost structure. There are three types of costs. First, the system incurs an inventory *holding cost* at a rate $h(X_t) \geq 0$. The assumptions on $h(\cdot)$ are the following.

Assumption I.2. *$h(\cdot)$ satisfies*

(i) $h(\cdot)$ is increasing convex;

(ii) $h(\cdot)$ is differentiable;

(iii) $h(0) = 0$; and

(iv) $h(\cdot)$ is polynomially bounded, i.e., there exist constants $A_i > 0$, $i = 1, 2$, and an integer $n \in \mathbb{N}^+$, such that $h(x) \leq A_1 + A_2 x^n$, for all x.

Second, whenever the state of the system is switched from $Y_{t^-} = 0$ to $Y_t = 1$, a *setup cost* $K > 0$ is incurred. Third, there is a *shortage cost* $c > 0$ for each unit of Z_t used to prevent the inventory level from dropping below 0, which is the amount of demand lost.

An *admissible policy* π is defined by a sequence of nonnegative stopping times $\tau_1 < \tau_2 < \tau_3 < \ldots$, a process Y_t, with $Y_0 = y$, and the corresponding switching probabilities at these points, such that

(i) $\tau_0 = 0$, $Y_{\tau_0} = y$ is the initial condition;

(ii) non-anticipating: for $n \geq 1$, $\tau_n \leq t$ is independent of $\{W_s - W_t, s > t\}$;

(iii) for $n \geq 1$, $P(Y_{\tau_n} = 1 - Y_{\tau_{n-1}}) > 0$; and

(iv) $P^{\pi}_{x,y}(\lim_{n\to\infty} \tau_n = \infty) = 1$.

Notice that condition (iii) allows the policy to be randomized, though most policies used in practice are stationary, with $P(Y_{\tau_n} = 1 - Y_{\tau_{n-1}}) = 1$. Let $\mathcal{A}$ be the class of all admissible policies.

For any policy $\pi \in \mathcal{A}$ and an initial system state (x, y), we define the total cost up to time T by

$$J^{\pi}_{x,y}(T) := E^{\pi}_{x,y}\left[\int_0^T h(X_t)dt + \sum_{0\le s\le T} K\ \delta^{+}(\Delta Y_s) + \int_0^T c\ dZ_t\right],$$

where $\Delta Y_s := Y_s - Y_{s^-} = Y_s - \lim_{t\to s^-} Y_t$, and $\delta^+(x) : \{-1, 0, 1\} \to \{0, 1\}$ is defined by $\delta^+(-1) = 0$, $\delta^+(0) = 0$, $\delta^+(1) = 1$. Finally, the *long-run average cost* is defined by

$$AC^{\pi}_{x,y} := \limsup_{T\to\infty} \frac{J^{\pi}_{x,y}(T)}{T}.$$

Let

$$AC_{x,y} := \inf_{\pi\in\mathcal{A}} AC^{\pi}_{x,y}.$$

An admissible policy π^* is called optimal if $AC^{\pi^*}_{x,y} = AC_{x,y}$ for all states (x, y). Our objective is to find an optimal policy for controlling the production/inventory system.

1.2.2 Lower bound for average cost.

In this subsection, we derive a lower bound for the average cost by virtue of the generalized Ito's formula. Roughly speaking, the lower bound provides a sufficient condition for optimality. If we can identify an admissible control policy whose relative value function satisfies this sufficient condition, then the average cost of this policy achieves the lower bound for the average cost among a large class of policies, thus it has to be optimal among this class.

The following result follows from an immediate application of the generalized Ito's formula for multi-dimensional stochastic processes, see e.g., the proof of Theorem 1 in Duckworth and Zervos (2001) [12], and for the single-dimensional case with jump-diffusions, (2.16) in Harrison (1983) [20]. Thus, its proof is omitted.

Proposition I.3. *Suppose that $f(x,y) : \mathbb{R} \times \{0,1\} \to \mathbb{R}$ is continuously differentiable, and has a continuous second derivative at all but a finite number of points with respect to x. Then for each time $T > 0$, initial state $x \in \mathbb{R}, y \in \{0,1\}$, and admissible policy π,*

$$\begin{aligned} E^{\pi}_{x,y}\left[f(X_T, Y_T)\right] = f(x,y) + E^{\pi}_{x,y}\Bigg[\int_0^T \Bigg\{ \left(\frac{1}{2}\sigma_0^2 f''(X_t,0) - \mu_0 f'(X_t,0)\right)(1-Y_t) + \\ + \left(\frac{1}{2}\sigma_1^2 f''(X_t,1) + \mu_1 f'(X_t,1)\right) Y_t \Bigg\} dt\Bigg] + E^{\pi}_{x,y}\left[\int_0^T f'(X_t,1) Y_t dZ_t\right] \\ + E^{\pi}_{x,y}\left[\int_0^T f'(X_t,0)(1-Y_t) dZ_t\right] + E^{\pi}_{x,y}\left[\sum_{0 \le s \le T} \Delta f(X_s, Y_s)\right], \end{aligned} \tag{1.1}$$

where f' and f'' are derivatives with respect to x, and $\Delta f(X_s, Y_s) = f(X_s, Y_s) - f(X_s, Y_{s^-})$.

In what follows, we use Γ_0 and Γ_1 to denote the infinitesimal generator associated with the two production modes, i.e.,

$$\begin{aligned} \Gamma_0 f(x,0) &:= -\mu_0 f'(x,0) + \frac{1}{2}\sigma_0^2 f''(x,0), \\ \Gamma_1 f(x,1) &:= \mu_1 f'(x,1) + \frac{1}{2}\sigma_1^2 f''(x,1). \end{aligned}$$

The important result below shows that, when these functions satisfy certain conditions, they can provide a lower bound on the optimal average cost.

Proposition I.4. *(Lower bound) Suppose the function $f(x,y)$ is polynomially bounded with respect to x and satisfies the conditions in Proposition I.3, and there is positive* *number γ such that the following conditions are satisfied:*

$$\Gamma_0 f(x,0) + h(x) - \gamma \geq 0, \tag{1.2}$$

$$\Gamma_1 f(x,1) + h(x) - \gamma \geq 0, \tag{1.3}$$

$$f(x,1) - f(x,0) \geq -K, \tag{1.4}$$

$$f(x,0) - f(x,1) \geq 0, \tag{1.5}$$

$$f'(0,1) + c \geq 0, \tag{1.6}$$

$$f'(0,0) + c \geq 0, \tag{1.7}$$

then γ is a lower bound of the average cost for all the policies in $\mathcal{A}_f$, i.e.,

$$AC^\pi = AC^\pi_{x,y} \geq \gamma, \qquad \forall \pi \in \mathcal{A}_f, \tag{1.8}$$

where $\mathcal{A}_f$ is defined by

$$\mathcal{A}_f := \left\{ \pi \in \mathcal{A} : \liminf_{T\to\infty} \frac{E^\pi_{x,y} f(X_T, Y_T)}{T} \leq 0, \forall x \in \mathbb{R}, \forall y \in \{0,1\} \right\}. \tag{1.9}$$

Proof. It follows from Proposition I.3 and conditions (1.2) to (1.7) that

$$E^\pi_{x,y}\left[f(X_T,Y_T)\right] \geq f(x,y) + E^\pi_{x,y}\left[\int_0^T (\gamma - h(X_t))\,dt - cZ_T + \sum_{0\leq s\leq T} (-K)\ \delta^+(\Delta Y_s)\right].$$

Because $\pi \in \mathcal{A}_f$, we have

$$AC^\pi_{x,y} = \limsup_{T\to\infty} \frac{J^\pi_{x,y}(T)}{T} \geq \gamma + \limsup_{T\to\infty} \frac{f(x,y)}{T} - \liminf_{T\to\infty} \frac{E^\pi_{x,y}\left[f(X_T,Y_T)\right]}{T} \geq \gamma.$$

This proves (1.8). ∎

Remark I.5. *This proposition claims that γ is a lower bound for the average cost among policies in $\mathcal{A}_f$, which is a subset of $\mathcal{A}$. In §1.4, we will discuss the class* *of policies in $\mathcal{A}_f$. In particular, we will define a subclass of policies in $\mathcal{A}_f$ that is independent of f, thus it is in $\mathcal{A}_f$ for any f, as long as f is polynomially bounded. Moreover, $\mathcal{A}_f$ includes all those policies that shut off the production automatically when the inventory level is higher than an arbitrarily large number M. This is clearly a very reasonable assumption, and it shows that $\mathcal{A}_f$ includes most control policies of practical interest.*

Remark I.6. *For inventory control problems with infinite capacity, i.e., impulse control, it can be shown that $\mathcal{A}_f$, under very mild conditions, contains all admissible policies. For example, when the cost rate function $h(\cdot)$ is polynomially bounded, this would be true. See, Ormeci et al. [30] for the linear holding and shortage cost case (their argument has been extended by Dai and Yao (2011) [23] to the case with polynomially bounded convex holding and shortage cost function). The approach used in Ormeci et al. [30] is that, for the infinite capacity model with cost function f of the optimal band policy, if it happens with a policy π that $\liminf_{T\to\infty} E^{\pi}_{x,y}[(f(X_T, Y_T)]/T > 0$, then the average cost for policy π must be infinity. This is shown by using the fact that, for the impulse control problem, the relative value function f is always Lipschitz continuous. This, however, is not true for the finite capacity case. In our case, if $h(\cdot)$ is polynomially bounded with highest degree n, then the value function f can be shown to be polynomially bounded with highest degree $n+1$ (see the appendix for analysis), so one degree higher than that of $h(\cdot)$. Thus when $\liminf_{T\to\infty} E^{\pi}_{x,y}[f(X_T, Y_T)]/T > 0$, it cannot be shown that the cost function for policy π is infinity. See §1.4 for an example on this.*

1.2.3 Analysis of (s, S) policy.

In this subsection we focus on a special class of policies: (s, S) policy, with $0 \leq s < S$, such that every time the inventory level reaches S, the machine is turned off, and every time the inventory level drops to s, the machine is turned on. For each policy within this class, we derive an algebraic expression for the average cost in terms of s and S. Then, we show the optimal parameters s and S are uniquely determined by an equation.

Theorem I.7. *For an (s, S) policy with $0 \leq s < S$, the average cost can be expressed as*

$$\gamma(s, S) \;=\; \frac{\int_s^S G(x)dx + K}{\int_s^S H(x)dx},$$

where

$$G(x) \;=\; \frac{2}{\sigma_0^2}\int_x^\infty e^{-\lambda_0(\xi - x)}h(\xi)d\xi + \frac{2}{\sigma_1^2}\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi + ce^{-\lambda_1 x}, \tag{1.10}$$

$$H(x) \;=\; m - \frac{1}{\mu_1}e^{-\lambda_1 x}, \tag{1.11}$$

$$m \;=\; \frac{1}{\mu_0} + \frac{1}{\mu_1}, \tag{1.12}$$

$$\lambda_i \;=\; \frac{2\mu_i}{\sigma_i^2}, \qquad i = 0, 1.$$

Proof. The stochastic process X_t operating under an (s, S) policy is a regenerative process if we define a cycle as follows. Suppose the inventory level starts from S, and the initial production state is 0, i.e., $X_0 = S, Y_0 = 0$. So at first, X_t evolves as a Brownian motion with drift $-\mu_0$ and variance parameter σ_0^2. Denote the hitting time of s by T_1, then $E_{S,0}[T_1] < \infty$. According to the (s, S) policy, Y_t is switched to 1 at $t = T_1$, after which, X_t evolves as a Brownian motion with drift μ_1 and and variance parameter σ_1^2. Suppose it takes another time T_2 for the process X_t to hit S. We call the total period including T_1 and T_2 a cycle, and the periods of T_1 and T_2 downward stage and upward stage, respectively. Due to the *regenerative structure* of the (s, S) policy, the long run average cost equals to the expected cost divided by the expected length over one cycle. See, e.g., [19, p. 86-89].

Now we compute the total cost over one cycle under this policy. For the expected holding cost incurred during the downward stage, we define

$$w_d(x) = E_{x,0}\left[\int_0^{T_1} h(X_t)dt\right], \qquad x \geq s,$$

where recall that T_1 is the hitting time of s, and x is the starting inventory level. It is known that $w_d(x)$ satisfies an ordinary differential equation with boundary conditions (see, e.g., Karlin and Taylor [26, §15.3 pages 192-193]):

$$\frac{\sigma_0^2}{2}w_d''(x) - \mu_0 w_d'(x) + h(x) = 0,\ w_d(s) = 0,\ \lim_{x\to\infty} e^{-\nu x} w_d(x) = 0,\ \forall \nu > 0.$$

The solution to this equation is

$$w_d(x) = \int_s^x \left(\frac{2}{\sigma_0^2}\int_u^\infty e^{-\lambda_0(\xi-u)}h(\xi)d\xi\right) du. \tag{1.13}$$

Also notice that $w_d(x)$ satisfies the differential equation

$$\Gamma_0 w_d(x) + h(x) = 0 \tag{1.14}$$

for all value of $x \geq 0$, not limited to $x \geq s$.

Similarly, for the expected holding cost over the upward stage, we define

$$w_u(x) = E_{x,1}\left[\int_0^{T_2} h(X_t)dt + c\int_0^{T_2} dZ_t\right], \qquad x \leq S.$$

where T_2 is the hitting time of S, with the starting inventory level x. During this stage, $dX_t = \mu_1 dt + \sigma_1 dW_t^1 + dZ_t$, and accordingly $w_u(x)$ obeys the differential equation

$$\frac{\sigma_1^2}{2}w_u''(x) + \mu_1 w_u'(x) + h(x) = 0,\ w_u(S) = 0,\ w_u'(0) = -c.$$

The solution to the ordinary differential equation is

$$w_u(x) = \int_x^S \left(\frac{2}{\sigma_1^2} \int_0^u e^{\lambda_1(\xi-u)} h(\xi) d\xi + ce^{-\lambda_1 u} \right) du. \tag{1.15}$$

Note that $w_u(x)$ also satisfies the differential equation

$$\Gamma_1 w_u(x) + h(x) = 0 \tag{1.16}$$

for all values of $x \geq 0$.

The expected duration of the downward stage initiated from x is $E_{x,0}[T_1] = (x - s)/\mu_0$. To compute $T(x) := E_{x,1}[T_2]$, we note that it satisfies the following ordinary differential equation:

$$\frac{\sigma_1^2}{2} T''(x) + \mu_1 T'(x) + 1 = 0,\ T'(0) = 0,\ T(S) = 0.$$

The solution of this differential equation is

$$T(x) = E_{x,1}(T_2) = \frac{S - x}{\mu_1} + \frac{\sigma_1^2}{2\mu_1^2} \left(e^{-\lambda_1 S} - e^{-\lambda_1 x} \right) = \frac{1}{\mu_1} \int_x^S (1 - e^{-\lambda_1 x'}) dx'.$$

Consequently, the average cost for the system operating under an (s, S) policy can be expressed as

$$\begin{aligned} c(s, S) &= \frac{w_d(S) + w_u(s) + K}{E_{S,0}(T_1) + E_{s,1}(T_2)} \\ &= \frac{\int_s^S G(x)dx + K}{m(S - s) + \frac{\sigma_1^2}{2\mu_1^2}\left(e^{-\lambda_1 S} - e^{-\lambda_1 s}\right)} \\ &= \frac{\int_s^S G(x)dx + K}{\int_s^S H(x)dx}, \end{aligned} \tag{1.17}$$

where $G(x)$, $H(x)$ and m are defined in equations (1.10)-(1.12). □

To establish the existence and uniqueness of the optimal choice of s and S, we need the following lemma. Its proof is given in the appendix.

Lemma I.8. *If the holding cost $h(x)$ satisfies Assumption I.2, then*

(i) $G(x)$ *is a strictly convex function, and*

(ii) $\lim_{x\to\infty} G(x) = \infty$.

Thus, $G(x) > 0$ is a convex function converging to infinity. It is easy to check that $H(x)$ is an increasing concave function with $H(0) = m - 1/\mu_1 = 1/\mu_0 > 0$, and $\lim_{x\to\infty} H(x) = m$.

We want to search, among the class of (s, S) policies, the policy that minimizes $c(s, S)$. To that end, we introduce an auxiliary function: For all $\gamma \geq 0$, let

$$\begin{aligned} \ell_\gamma(s, S) &= \int_s^S G(x)dx - \gamma \int_s^S H(x)dx + K \\ &= \int_s^S \Big(G(x) - \gamma H(x) \Big) dx + K. \end{aligned}$$

For a fixed $\gamma > 0$, since $G(x) - \gamma H(x)$, defined on $x \geq 0$, is strictly convex and tends to infinity as $x \to \infty$, it has a unique minimum on $[0, \infty)$. Let this minimum be denoted by y^*_γ.

The following result is easy to prove so its proof is omitted.

Lemma I.9. (s^*, S^*) *minimizes* $c(s, S)$ *if and only if there exists* γ^* *such that*

$$\min_{0\leq s\leq S} \ell_{\gamma^*}(s, S) = \ell_{\gamma^*}(s^*, S^*) = 0.$$

Therefore, in what follows we first minimize $\min_{0\leq s\leq S} \ell_\gamma(s, S)$ for a given γ, and then we search for γ that satisfies Lemma I.9. The following lemma is useful in that regard. The proof is easy so is omitted.

Lemma I.10. *For any $\gamma \geq 0$, $\ell_\gamma(s,S)$ is increasing in S if and only if $G(S) - \gamma H(S) \geq 0$, and it is increasing in s if and only if $G(s) - \gamma H(s) \leq 0$.*

Let $(s(\gamma), S(\gamma))$ be the optimal solution of $\min_{s \leq S} \ell_\gamma(s,S)$. It is clear that $\{x \geq 0; G(x) - \gamma H(x) \leq 0\}$ is a null set if and only if γ is smaller than a positive critical value $\underline{\gamma}$, where $\underline{\gamma}$ is the smallest value of γ for which the two curves $\gamma H(x)$ and $G(x)$ touch each other. In particular, if $\gamma < \underline{\gamma}$, then $\ell_\gamma(s,S)$ in Lemma I.9 is positive for any $s < S$, i.e., any $\gamma < \underline{\gamma}$ cannot be achieved by an (s,S) policy. For convenience, if $\gamma < \underline{\gamma}$, then we let $s(\gamma) = S(\gamma) = y^*_\gamma$. Here and below, if not otherwise stated, we restrict our attention to those values of $\gamma \geq \underline{\gamma}$. It follows from Lemma I.10 that

$$\begin{aligned} s(\gamma) &= \min\left\{0 \leq x \leq y^*_\gamma;\ \ G(x) - \gamma H(x) \leq 0\right\}, && (1.18) \\ S(\gamma) &= \max\left\{x \geq y^*_\gamma;\ \ G(x) - \gamma H(x) \leq 0\right\}. && (1.19) \end{aligned}$$

As a result, the optimal $s(\gamma) \leq S(\gamma)$ exist for all γ, though they may be equal to each other. It is seen from this definition that $G(x) - \gamma H(x) \leq 0$ on $s(\gamma) \leq x \leq S(\gamma)$; and $G(x) - \gamma H(x) \geq 0$ on $0 \leq x \leq s(\gamma)$ and $x \geq S(\gamma)$. Because $G(x) - \gamma H(x)$ is strictly decreasing in γ, it can be seen that

$$A(\gamma) := \int_{s(\gamma)}^{S(\gamma)} \Big(G(x) - \gamma H(x)\Big) dx \qquad (1.20)$$

is strictly decreasing (and concave) in γ. $A(\gamma)$ is concave since

$$A(\gamma) = \min_{s \leq S} \int_s^S G(x) - \gamma \int_s^S H(x) dx$$

is the minimum of a family of concave functions of γ. In addition, $S(\gamma)$ is strictly increasing in γ and $s(\gamma)$ is non-increasing in γ.

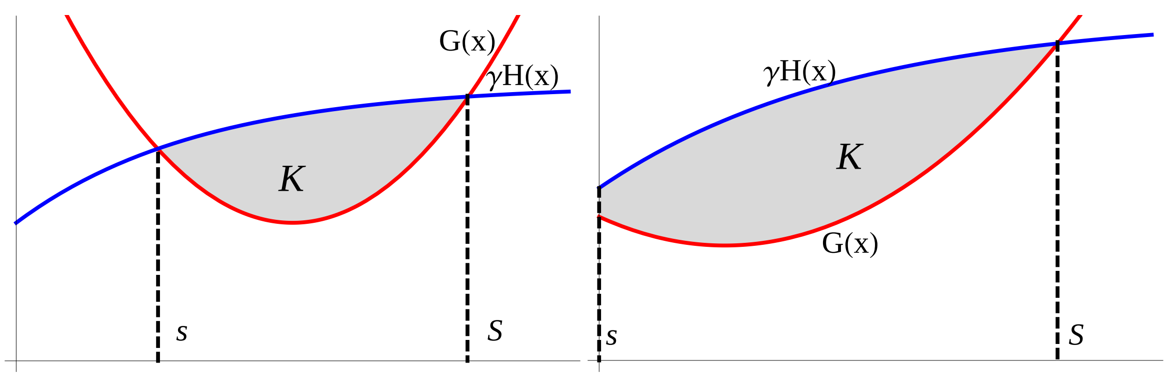
The following theorem presents the condition for parameters s and S to minimize the average cost function $c(s,S)$.

Theorem I.11. *The unique optimal s^* and S^* that minimize the average cost $c(s,S)$ is $s^* = s(\gamma^*)$ and $S^* = S(\gamma^*)$, where γ^* is uniquely determined by*

$$\int_{s(\gamma)}^{S(\gamma)} \Big(G(x) - \gamma H(x)\Big) dx = -K. \tag{1.21}$$

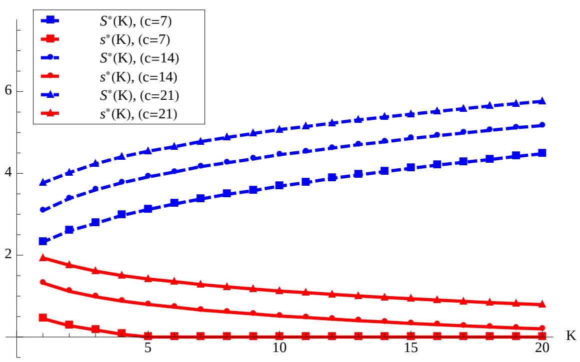
Proof. When $\gamma > \underline{\gamma}$, $A(\gamma)$ is strictly decreasing and tends to $-\infty$ as $\gamma \to \infty$. If $\gamma = 0$, then it is easily seen by $G(x) \geq 0$ that $s(0) = S(0) = y_0^*$ and $A(0) = 0$. Thus, by continuity of $A(\gamma)$, there exists a unique γ^* that satisfies $A(\gamma^*) = -K$, or (1.21), or $\ell_{\gamma^*}(s(\gamma^*), S(\gamma^*)) = 0$. The optimality of $s^* = s(\gamma^*)$ and $S^* = S(\gamma^*)$ follows from Lemma I.9. The uniqueness of s^* and S^* are easy to show due to the convexity of $G(x) - \gamma^* H(x)$. $\square$

An illustration of the optimal $(s(\gamma), S(\gamma))$ is given in Figure 1.1. From Figure 1.1, it is easily seen that as $K \to \infty$, S^* increases, and s^* decreases. At the same time, the value of γ^* increases. To further illustrate the effect of parameters K and c on the optimal band policy, we conduct a numerical analysis as follows: Let $h(x) = 2x$, $\mu_0 = \mu_1 = 0.5$, and consider two sets of values for σ_0 and σ_1: $\sigma_0 = \sigma_1 = 2$ and $\sigma_0 = 2$, $\sigma_1 = 3$. The optimal average cost γ^*, together with the optimal s^*, S^* as functions of K for different values of c are summarized in Figure 1.2.

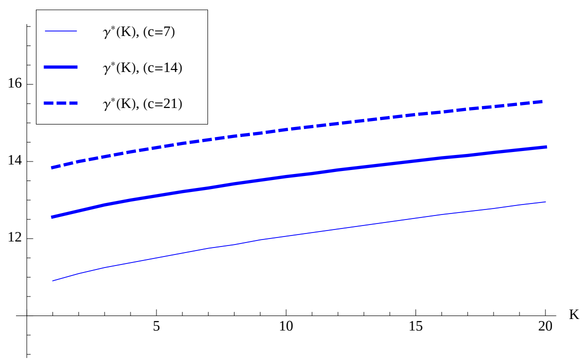


(a) $s > 0$ (b) $s = 0$

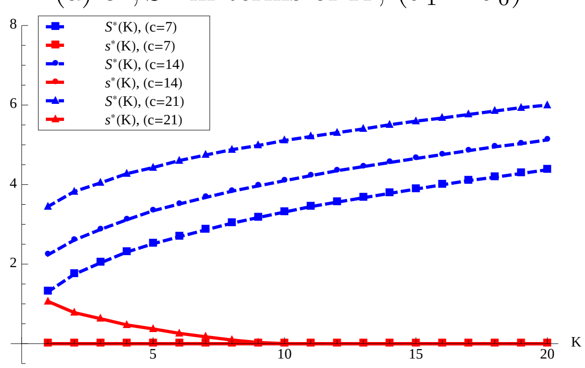
Figure 1.1: Optimal choice of s and S (lost-sales case)



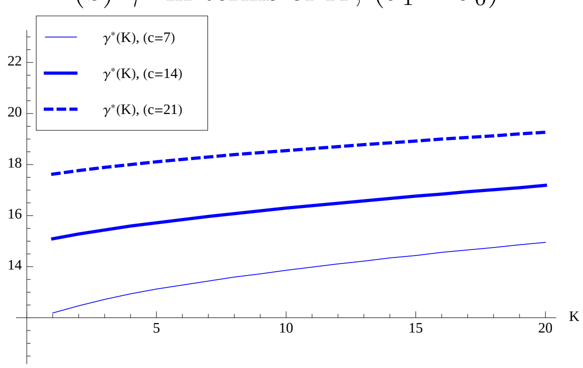
(a) s^*, S^* in terms of K, ($\sigma_1 = \sigma_0$)



(b) γ^* in terms of K, ($\sigma_1 = \sigma_0$)



(c) s^*, S^* in terms of K, ($\sigma_1 > \sigma_0$)



(d) γ^* in terms of K, ($\sigma_1 > \sigma_0$)

Figure 1.2: The effect of K, c on the average cost and optimal choice of s, S

1.2.4 Optimal policy.

In the previous subsection, we have identified a policy (s^*, S^*), which is optimal among the class of (s, S) policies. In this subsection, we aim to find the optimal policy within a larger class of policies.

The first question is whether it is possible for a non- (s, S) type of policy to be optimal. The answer is affirmative. In the following, we show that, within a very large class of policies, the optimal one is either to never turn on the machine for production, or produce according to an (s, S) policy. We shall give the range of system parameters within which each of these two policies is optimal.

If we never turn on the machine for production, then the stochastic process under consideration is a regulated Brownian motion with drift $-\mu_0 < 0$ and variance parameter σ_0^2, and a reflection boundary at 0. Because the machine is never turned on, there is no setup cost and there is only the holding and the shortage cost. The average cost for this process is readily computed, and it is given by (see e.g., Harrison [19, §5.6])

$$\gamma_0 = G(0)/H(0) = \frac{2\mu_0}{\sigma_0^2} \int_0^\infty e^{-\lambda_0 \xi} h(\xi) d\xi + \mu_0 c. \tag{1.22}$$

Since $G(x) - \gamma_0 H(x)$ is equal to 0 when $x = 0$, we want to identify the other zero point for this convex function. To that end, let $S_0(\gamma_0)$ be the maximum zero point for $G(x) - \gamma_0 H(x)$, i.e.,

$$S_0(\gamma_0) = \max\{x \geq 0;\ G(x) - \gamma_0 H(x) \leq 0\}. \tag{1.23}$$

For convenience, in what follows we simply write $S_0(\gamma_0)$ as S_0.

By checking the sign of the derivative of $G(x) - \gamma_0 H(x)$ at $x = 0$, the following result can be easily established.

Lemma I.12. *$S_0 > 0$ if and only if the system parameters satisfy*

$$c > c_0 := \frac{\lambda_0}{\mu_0 + \mu_1}\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right)\int_0^\infty e^{-\lambda_0 \xi} h(\xi) d\xi. \tag{1.24}$$

Note that this condition is always satisfied when $\sigma_1 \leq \sigma_0$. If (1.24) is satisfied, then we are interested in the system parameters that additionally satisfy

$$\int_0^{S_0} \left(G(x) - \gamma_0 H(x)\right) dx < -K, \tag{1.25}$$

Remark I.13. *Note that conditions (1.24) and (1.25) can be combined into only (1.25), since when condition (1.24) is not satisfied, then $S_0 = 0$ and (1.25) will not be satisfied. Nevertheless, since the region of parameters satisfying (1.25) are determined by two inequalities (1.24) and (1.25), we shall refer to both (1.24) and (1.25).*

We first consider the case when the system parameters satisfy both (1.24) and (1.25). We will prove that, in this case we can identify an (s^*, S^*) policy, with parameters determined in Theorem 2.2, that is optimal within a large class of policies.

We first prove that, under conditions (1.24) and (1.25), the s^* and S^* in Theorem I.11 satisfy $0 < s^* < S^*$. Using the fact that $\ell_\gamma(s(\gamma), S(\gamma))$ is strictly decreasing in γ when $A(\gamma) < 0$ and that γ^* is determined by $\ell_{\gamma^*}(s(\gamma^*), S(\gamma^*)) = 0$, we conclude by (1.25) that $\gamma_0 > \gamma^*$. Then, it follows from $S(\gamma)$ is strictly increasing and $s(\gamma)$ non-increasing in γ, that

$$S(\gamma_0) > S(\gamma^*) = S^* \geq s^* = s(\gamma^*) \geq s(\gamma_0) = 0.$$

Because $G(0) - \gamma^* H(0) < G(0) - \gamma_0 H(0) = 0$, thereby we have $s^* = s(\gamma^*) > s(\gamma_0) = 0$ and $0 < s^* < S^*$. Hence, it follows from $S^* \geq s^* > 0$ that the minimizer of

$G(x) - \gamma^* H(x)$ is positive, and that $G'(x) - \gamma^* H'(x) \le 0$ on $0 \le x \le s^*$ and $G'(x) - \gamma^* H'(x) \ge 0$ on $x \ge S^*$.

Using policy (s^*, S^*) and the corresponding γ^*, we define value functions $v(x, y)$ as follows:

$$v(x,0) = \begin{cases} w_d(x) - \gamma^* \left(\frac{x - s^*}{\mu_0}\right) + v(s^*, 1) + K, & x > s^*, \\ v(x,1) + K, & x \le s^*, \end{cases} \tag{1.26}$$

$$v(x,1) = \begin{cases} w_u(x) - \gamma^* \left[\frac{S^* - x}{\mu_1} + \frac{\sigma_1^2}{2\mu_1^2}\left(e^{-\lambda_1 S^*} - e^{-\lambda_1 x}\right)\right], & x < S^*, \\ v(x,0), & x \ge S^*, \end{cases} \tag{1.27}$$

where w_u and w_d are defined by (1.15) and (1.13) using s^* and S^*. The value function $v(x, y)$ can be interpreted as the cost incurred starting from the current state (x, y) until the end of the cycle, i.e., the process reaches $(S^*, 1)$, minus the expected remaining time of the cycle multiplied by γ^*. It is easy to verify that both $v(x, 0)$ and $v(x, 1)$ are continuous functions of x. We only prove the continuity of $v(x, 1)$ at S^*. Since $\int_{s^*}^{S^*} (G(x) - \gamma^* H(x))\,dx + K = 0$, we have

$$\begin{aligned} v(S^*, 1) = v(s^*, 0) &= w_d(S^*) - \gamma^* \left(\frac{S^* - s^*}{\mu_0}\right) + v(s^*, 1) + K \\ &= w_d(S^*) + w_u(s^*) - \gamma^* \left(\frac{S^* - s^*}{\mu_0}\right) - \gamma^* \left[\frac{S^* - s^*}{\mu_1} + \frac{\sigma_1^2}{2\mu_1}\left(e^{-\lambda_1 S^*} - e^{-\lambda_1 s^*}\right)\right] + K \\ &= \int_{s^*}^{S^*} (G(x) - \gamma^* H(x))\,dx + K \\ &= 0 \\ &= \lim_{x \to S^{*-}} v(x, 1). \end{aligned}$$

Thus $v(x, 1)$ is continuous at $x = S^*$.

The next theorem states that, when the system parameters satisfy both (1.24) and (1.25), then an (s^*, S^*) policy is optimal among the class $\mathcal{A}_v$,

Theorem I.14. *Suppose the system parameters satisfy both (1.24) and (1.25), then* *the (s^*, S^*) policy is optimal among all policies in $\mathcal{A}_v$.*

Proof. It suffices to verify that the relative value function $v(x, y)$ defined above satisfies all the conditions in Proposition I.4.

For the condition (1.6) on the derivative of $v(x, y)$ at $x = 0$, by (1.27) we have

$$v'(0, 1) = w_u'(0) + \frac{\gamma^*}{\mu_1} - \frac{\gamma^*}{\mu_1} e^{-\lambda_1 0} = -c,$$

hence inequality (1.6) is satisfied. For $v(x, 0)$, since for this case we have $s^* > 0$, thus $v'(0, 0) = v'(0, 1) = -c$, thereby inequality (1.7) is also satisfied.

From the definitions of w_u, w_d, and Γ_y, it is easy to see that when $x > s^*$, $\Gamma_0 v(x, 0) + h(x) - \gamma^* = 0$ and when $x < S^*$, $\Gamma_1 v(x, 1) + h(x) - \gamma^* = 0$. Thus, to complete the proof of (1.2) and (1.3), we need to verify $\Gamma_0 v(x, 0) + h(x) - \gamma^* \geq 0$ on $x \leq s^*$ and $\Gamma_1 v(x, 1) + h(x) - \gamma^* \geq 0$ on $x \geq S^*$.

Suppose $x \leq s^*$. By the definition of $v(x, 0)$ on this range and (1.13), $\Gamma_0 v(x, 0) + h(x) - \gamma^* \geq 0$ is equivalent to

$$\Gamma_0 v(x, 1) - \Gamma_0 w_d(x) - \gamma^* \geq 0.$$

Substituting (1.27) into the equation yields

$$\Gamma_0 \left(w_u(x) + \frac{\gamma^* x}{\mu_1} + \frac{\sigma_1^2}{2\mu_1^2} \gamma^* e^{-\lambda_1 x} \right) - \Gamma_0 w_d(x) - \gamma^* \geq 0.$$

This can further be simplified as

$$\begin{aligned}
& -\Gamma_0 \left(w_d(x) - w_u(x) \right) + \Gamma_0 \left(\frac{\gamma^* x}{\mu_1} + \frac{\sigma_1^2}{2\mu_1^2} \gamma^* e^{-\lambda_1 x} \right) - \gamma^* \\
= & -\frac{\sigma_0^2}{2} G'(x) + \mu_0 G(x) - \frac{\gamma^* \mu_0}{\mu_1} + \frac{\gamma^* \sigma_0^2}{\sigma_1^2} e^{-\lambda_1 x} + \frac{\gamma^* \mu_0}{\mu_1} e^{-\lambda_1 x} - \gamma^* \\
= & -\frac{\sigma_0^2}{2} \left(G'(x) - \frac{2\gamma^*}{\sigma_1^2} e^{-\lambda_1 x} \right) + \mu_0 \left(G(x) - \frac{\gamma^*}{\mu_0} - \frac{\gamma^*}{\mu_1} + \frac{\gamma^*}{\mu_1} e^{-\lambda_1 x} \right).
\end{aligned}$$

Since $S^* \geq s^* > 0$, it holds that on $x \leq s^*$,

$$\begin{aligned} G(x) \geq \gamma^* H(x) &= \frac{\gamma^*}{\mu_0} + \frac{\gamma^*}{\mu_1} - \frac{\gamma^*}{\mu_1} e^{-\lambda_1 x}, \\ G'(x) \leq \gamma^* H'(x) &= \frac{2\gamma^*}{\sigma_1^2} e^{-\lambda_1 x}. \end{aligned}$$

Therefore, it leads to, whenever $x \leq s^*$,

$$\Gamma_0 v(x, 0) + h(x) - \gamma^* \geq 0.$$

Next, we verify $\Gamma_1 v(x, 1) + h(x) - \gamma^* \geq 0$ on $x \geq S^*$, which is the same as

$$\Gamma_1 v(x, 0) - \Gamma_1 w_u(x) - \gamma^* \geq 0.$$

Substituting (1.26) yields

$$\Gamma_1 \left(w_d(x) - \frac{\gamma^* x}{\mu_0} \right) - \Gamma_1 w_u(x) - \gamma^* \geq 0.$$

This can be simplified as

$$\begin{aligned} &\Gamma_1 \left(w_d(x) - w_u(x) \right) - \Gamma_1 \left(\frac{\gamma^* x}{\mu_0} \right) - \gamma^* \\ =& \frac{\sigma_1^2}{2} G'(x) + \mu_1 G(x) - \frac{\gamma^* \mu_1}{\mu_0} - \gamma^* \\ =& \frac{\sigma_1^2}{2} \left(G'(x) - \frac{2\gamma^*}{\sigma_1^2} e^{-\lambda_1 x} \right) + \mu_1 \left(G(x) - \frac{\gamma^*}{\mu_0} - \frac{\gamma^*}{\mu_1} + \frac{\gamma^*}{\mu_1} e^{-\lambda_1 x} \right). \end{aligned}$$

By the definition of S^*, we have, on $x \geq S^*$,

$$\begin{aligned} G(x) \geq \gamma^* H(x) &= \frac{\gamma^*}{\mu_0} + \frac{\gamma^*}{\mu_1} - \frac{\gamma^*}{\mu_1} e^{-\lambda_1 x}, \\ G'(x) \geq \gamma^* H'(x) &= \frac{2\gamma^*}{\sigma_1^2} e^{-\lambda_1 x}. \end{aligned}$$

This shows that $\Gamma_1 v(x,1) + h(x) - \gamma^* \geq 0$ for all $x \geq S^*$. Therefore, inequalities (1.2)-(1.3) have been proved.

Finally, we prove $v(x,y)$ satisfies conditions (1.4)-(1.5). By their definitions, the inequalities are clearly satisfied on $x \geq S^*$ and $x \leq s^*$. If $s^* < x < S^*$, then

$$\frac{d}{dx}\Big(v(x,0) - v(x,1)\Big) = G(x) - \gamma^* H(x) \leq 0. \tag{1.28}$$

Thus

$$\begin{aligned}
[v(x,0) - v(x,1)] = & \ [v(s^*,0) - v(s^*,1)] + \int_{s^*}^{x} \Big(G(u) - \gamma^* H(u)\Big) du \\
= & \ K + \int_{s^*}^{x} \Big(G(u) - \gamma^* H(u)\Big) du
\end{aligned}$$

On the other hand, when $s^* < x < S^*$, we have

$$-K \leq \int_{s^*}^{x} \Big(G(u) - \gamma^* H(u)\Big) du \leq 0,$$

hence we obtain

$$0 \leq v(x,0) - v(x,1) \leq K, \ \forall x \in [s^*, S^*].$$

This shows that (1.4)-(1.5) hold for all x.

We now verify that $v(x,y)$ have continuous first order derivatives in x. From their definitions, this is clearly true when $x \neq s^*, S^*$, hence we only need to verify the continuity at these two points. Here we only verify the continuity of $v'(x,0)$ at point $x = s^*$ since the verification of continuity of $v'(x,1)$ at $x = S^*$ is similar. The optimality condition

$$\gamma^* H(s^*) - G(s^*) = 0$$

implies that

$$
\begin{aligned}
&\lim_{x\to(s^*)^+} v'(x,0) - \lim_{x\to(s^*)^-} v'(x,0) \\
&= [w_d'(s^*) - \gamma^*/\mu_0] - v'(s^*,1) \\
&= [w_d'(s^*) - \gamma^*/\mu_0] - \left[w_u'(s^*) + \gamma^*/\mu_1 - (\gamma^*/\mu_1)\, e^{-\lambda_1 s^*}\right] \\
&= G(s^*) - \gamma^* H(s^*) \\
&= 0,
\end{aligned}
$$

thereby $v'(x,0)$ is continuous at s^*.

To summarize, we have shown that all the conditions (1.2)-(1.7) are satisfied by $v(x,y)$; and the continuity conditions are also verified. By the definition of $v(x,y)$, it is clear that its second derivative is continuous at all but a finite number of points, i.e., possibly not continuous at s^* and S^*. Therefore, it follows from Proposition I.4 that γ^* is an achievable lower bound on the long-run average cost for the policies in the set $\mathcal{A}_v$, implying that this (s^*, S^*) policy is optimal in $\mathcal{A}_v$. □

The theorem above shows that, when the system parameters satisfy both (1.24) and (1.25), then the optimal policy is (s^*, S^*), which we have computed in Theorem 2.2. What happens if the system parameters do not satisfy any of them? The following theorem shows that in that case, the "never-turn-on-the-machine" policy is optimal, again within a large class of policies. This implies that, in such range of cost parameters, it is not economically justified for the firm to enter the business. Recall that γ_0 and S_0 are defined in (1.22) and (1.23).

Theorem I.15. *If the system parameters either do not satisfy (1.24), or they satisfy (1.24) but do not satisfy (1.25), then the "never turn on the machine" policy is optimal* *within the class of policies $\mathcal{A}_g$, where*

$$g(x,0) = \int_0^x \left(\frac{2}{\sigma_0^2} \int_u^\infty e^{-\lambda_0(\xi-u)} h(\xi) d\xi \right) du - \frac{\gamma_0 x}{\mu_0}, \tag{1.29}$$

$$g(x,1) = \begin{cases} g(x,0), & x \geq S_0, \\ \\ \int_x^{S_0} \left(\frac{2}{\sigma_1^2} \int_0^u e^{\lambda_1(\xi-u)} h(\xi) d\xi + c e^{-\lambda_1 u} \right) du + \frac{\gamma_0 x}{\mu_1} + \frac{\gamma_0}{\lambda_1 \mu_1} e^{-\lambda_1 x} + \theta, & 0 \leq x < S_0, \end{cases} \tag{1.30}$$

where θ is a constant given by

$$\theta = g(S_0, 0) - \frac{\gamma_0 S_0}{\mu_1} - \frac{\gamma_0}{\lambda_1 \mu_1} e^{-\lambda_1 S_0}.$$

Proof. We first prove the result for the case when (1.24) is satisfied but (1.25) is not satisfied. In this case, $S_0 > 0$. Since (1.20) is strictly decreasing in γ, and γ^* satisfies (1.21), it follows that in this case we have $\gamma_0 \leq \gamma^*$. Note that γ_0 is the average cost for the "never-turn-on-the-machine" policy, while γ^* is that of the best (s, S) policy. This shows that the "never-turn-on-the-machine" policy is better than the best (s, S) policy. In the following, we prove that this policy is optimal within the larger class of policies, $\mathcal{A}_g$, but using Proposition I.4.

It is easy to verify, using the properties of (1.13) and (1.15), that $g(x, y)$ satisfies the differential equation

$$\Gamma_0 g(x,0) + h(x) - \gamma_0 = 0, \qquad \text{for all } x$$

and

$$\Gamma_1 g(x,1) + h(x) - \gamma_0 = 0, \ \forall \ 0 \leq x < S_0.$$

To prove that $g(x,y)$ satisfy (1.2) and (1.3), we need to verify $g(x,1)$ satisfies $\Gamma_1 g(x,1)+h(x)-\gamma_0 \geq 0$ on $x \geq S_0$. To that end, note that, by (1.15), the function defined by

$$\tilde{g}(x) := \int_x^{S_0} \left(\frac{2}{\sigma_1^2} \int_0^u e^{\lambda_1(\xi-u)} h(\xi) d\xi + c e^{-\lambda_1 u} \right) du + \frac{\gamma_0 x}{\mu_1}$$

satisfies the differential equation

$$\Gamma_1 \tilde{g}(x) + h(x) - \gamma_0 = 0, \ \forall \ x \geq 0.$$

So it suffices to prove

$$\Gamma_1 g(x,1) - \Gamma_1 \tilde{g}(x) = \Gamma_1 g(x,0) - \Gamma_1 \tilde{g}(x) \geq 0, \ \forall \ x \geq S_0.$$

We have

$$\begin{aligned}
&\Gamma_1 g(x,0) - \Gamma_1 \tilde{g}(x) \\
= {}& \mu_1 \left(\frac{2}{\sigma_0^2} \int_x^\infty e^{-\lambda_0(\xi-x)} h(\xi) d\xi - \frac{\gamma_0}{\mu_0} \right) \\
&+ \frac{\sigma_1^2}{2} \left[\frac{2}{\sigma_0^2} \left(-h(x) + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi-x)} h(\xi) d\xi \right) \right] \\
&- \mu_1 \left(-\frac{2}{\sigma_1^2} \int_0^x e^{\lambda_1(\xi-x)} h(\xi) d\xi - c e^{-\lambda_1 x} + \frac{\gamma_0}{\mu_1} \right) \\
&- \frac{\sigma_1^2}{2} \left[-\frac{2}{\sigma_1^2} \left(h(x) - \lambda_1 \int_0^x e^{\lambda_1(\xi-x)} h(\xi) d\xi \right) + c\lambda_1 e^{-\lambda_1 x} \right].
\end{aligned} \tag{1.31}$$

By the definition of S_0, on $x \geq S_0$ we have $G'(x) - \gamma_0 H'(x) \geq 0$ and $G(x) - \gamma_0 H(x) \geq 0$. These imply the following inequalities

$$\begin{aligned}
&\frac{2}{\sigma_0^2} \left(-h(x) + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi-x)} h(\xi) d\xi \right) \\
&\qquad + \frac{2}{\sigma_1^2} \left(h(x) - \lambda_1 \int_0^x e^{\lambda_1(\xi-x)} h(\xi) d\xi \right) - c\lambda_1 e^{-\lambda_1 x} - \gamma_0 \left(\frac{\lambda_1}{\mu_1} e^{-\lambda_1 x} \right) \ \geq \ 0,
\end{aligned}$$

$$\frac{2}{\sigma_0^2}\int_x^\infty e^{-\lambda_0(\xi-x)}h(\xi)d\xi + \frac{2}{\sigma_1^2}\int_0^x e^{\lambda_1(\xi-x)}h(\xi)d\xi + ce^{-\lambda_1 x} - \gamma_0\left(m - \frac{1}{\mu_1}e^{-\lambda_1 x}\right) \ \geq \ 0.$$

Substituting these two inequalities into (1.31) yields

$$\Gamma_1 g(x,0) - \Gamma_1\tilde{g}(x) \geq \mu_1\gamma_0\left(m - \frac{1}{\mu_1}e^{-\lambda_1 x}\right) + \frac{\sigma_1^2}{2}\gamma_0\left(\frac{\lambda_1}{\mu_1}e^{-\lambda_1 x}\right) - \frac{\mu_1\gamma_0}{\mu_0} - \gamma_0 = 0.$$

Thus, (1.2) and (1.3) are shown to be satisfied.

We next prove (1.4) and (1.5). By their definitions, it is easy to see that $g(x,y)$ is continuous in x for $y = 0,1$, and $g(x,1) - g(x,0) = 0$ on $x \geq S_0$. The differentiability of $g(x,1)$ at $x = S_0$ can be shown easily due to $G(S_0) - \gamma_0 H(S_0) = 0$ so is omitted here. For any $x \in [0, S_0]$, we have

$$\begin{aligned}
g(x,1) - g(x,0) &= g(S_0,1) - g(S_0,0) - \int_x^{S_0}(g(u,1) - g(u,0))'\,du \\
&= -\int_x^{S_0}(g(u,1) - g(u,0))'du \\
&= -\int_x^{S_0}\Bigg[-\frac{2}{\sigma_1^2}\int_0^u e^{\lambda_1(\xi-u)}h(\xi)d\xi - ce^{-\lambda_1 u} + \frac{\gamma_0}{\mu_1} - \frac{\gamma_0}{\mu_1}e^{-\lambda_1 u} \\
&\qquad\qquad - \frac{2}{\sigma_0^2}\int_u^\infty e^{-\lambda_0(\xi-u)}h(\xi)d\xi + \frac{\gamma_0}{\mu_0}\Bigg]du \\
&= \int_x^{S_0}\Big(G(u) - \gamma_0 H(u)\Big)du \\
&\geq \int_0^{S_0}\Big(G(u) - \gamma_0 H(u)\Big)du \\
&\geq -K,
\end{aligned}$$

where the first inequality follows from $G(u) - \gamma_0 H(u) \leq 0$ on $0 \leq u \leq S_0$, and the last inequality follows from the opposite of (1.25). The last equality above also shows, again by $G(u) - \gamma_0 H(u) \leq 0$ on $0 \leq u \leq S_0$, that $g(x,1) - g(x,0) \leq 0$. Thus (1.4) and (1.5) are proved for $g(x,y)$.

We finally prove (1.6) and (1.7). By the definition of γ_0, we have

$$g'(0,0) = \frac{2}{\sigma_0^2}\int_0^\infty e^{-\lambda_0 \xi}h(\xi)d\xi - \frac{\gamma_0}{\mu_0} = -c.$$

For $g'(0,1)$, since $S_0 > 0$, it holds that, for $0 \le x < S_0$,

$$g'(x,1) = -\frac{2}{\sigma_1^2}\int_0^x e^{\lambda_1(\xi - x)}h(\xi)d\xi - ce^{-\lambda_1 x} + \frac{\gamma_0}{\mu_1} - \frac{\gamma_0}{\mu_1}e^{-\lambda_1 x},$$

thereby $g'(0,1) = -c$. Hence (1.6) and (1.7) are also verified. This proves the result for the case when (1.24) holds but (1.25) is not satisfied.

Next, we consider the case when (1.24) is not satisfied. Then $S_0 = 0$. By the definitions of $s(\gamma)$ and $S(\gamma)$, we also have $s(\gamma_0) = S(\gamma_0) = 0$ and as a result, $\ell_{\gamma_0}(s(\gamma_0), S(\gamma_0)) = K$. Since $\ell_\gamma(s(\gamma), S(\gamma))$ is strictly decreasing in γ, it follows that the optimal γ^*, determined by $\ell_{\gamma^*}(s(\gamma^*), S(\gamma^*)) = 0$, satisfies $\gamma_0 < \gamma^*$. As γ_0 is the average cost of the "never-turn-on-the-machine" policy, while γ^* is minimum average cost among the class of (s, S) policies, this shows that "never-turn-on-the-machine" policy is also better than any of the (s, S) policy in this case, and in the following we use Proposition I.4 to prove that the "never-turn-on-machine" is optimal among all policies in $\mathcal{A}_g$, where $g(x, y)$ is still as defined in the theorem.

In this case, $g(x, 0) = g(x, 1)$ for all $x \ge 0$, and again, we need to show that this function satisfies all the conditions of Proposition I.4, (1.2)-(1.7), and that the lower bound is achieved by using the "never-turn-on-machine" policy.

Since the first part of (1.29) is a special case of (1.13), which satisfies (1.14), we conclude that $g(x, y)$ satisfies

$$\Gamma_0 g(x,0) + h(x) - \gamma_0 = 0.$$

Thus (1.2) is satisfied. The second condition, (1.3), can be written as

$$\mu_1 \left(\frac{2}{\sigma_0^2} \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi - \frac{\gamma_0}{\mu_0} \right) \\ + \frac{\sigma_1^2}{2} \left[\frac{2}{\sigma_0^2} \left(-h(x) + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi \right) \right] + h(x) - \gamma_0 \geq 0, \ \forall\, x \geq 0. \quad (1.32)$$

The third term on the left hand side, $h(x)$, is increasing in x. We now prove that the first two terms on the left hand side are also increasing in x. The derivative of the first term, if we ignore the constant positive coefficient, is

$$\begin{aligned} & -h(x) + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi \\ \geq & -h(x) + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} h(x) d\xi \\ = & \ 0, \end{aligned}$$

where inequality follows from $h(\xi) \geq h(x)$ on $\xi \geq x$. For the second term, we note that

$$\begin{aligned} & -h(x) + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi \\ = & -\lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} h(x) d\xi + \lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi \\ = & \ \lambda_0 \int_x^\infty e^{-\lambda_0(\xi - x)} \left(h(\xi) - h(x) \right) d\xi \\ = & \ \lambda_0 \int_0^\infty e^{-\lambda_0 y} \left(h(x + y) - h(x) \right) dy. \end{aligned}$$

Since $h(\cdot)$ is a convex function, for any fixed y, $h(x+y) - h(x)$ is increasing in x, thus the integral above is also increasing in x.

Thus, the left hand side of (1.32) is increasing in x, and to prove (1.32), it suffices to prove it for $x = 0$. For notational convenience, let

$$\delta = \int_0^\infty e^{-\lambda_0 \xi} h(\xi) d\xi,$$

then $\gamma_0 = \lambda_0 \delta + c\mu_0$, and (1.32) becomes

$$\frac{2\mu_1}{\sigma_0^2}\delta - \frac{\gamma_0 \mu_1}{\mu_0} + \frac{\sigma_1^2 \lambda_0}{\sigma_0^2}\delta - \gamma_0 \geq 0.$$

Substituting γ_0 into the left hand side, it is simplified to

$$\frac{1}{\mu_0 + \mu_1}\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right)\lambda_0 \delta \geq c.$$

This is precisely the range of parameters we are considering, i.e., the opposite of (1.24), thus it ought to be satisfied. This proves (1.3).

Conditions (1.4) and (1.5) are obviously satisfied since $g(x, 0) = g(x, 1)$. Moreover, from the definition of γ_0, we have

$$g'(0, y) = \frac{2}{\sigma_0^2}\int_0^\infty e^{-\lambda_0(\xi - u)} h(\xi) d\xi - \frac{\gamma_0}{\mu_0} = -c.$$

This proves (1.6) and (1.7).

Therefore, we have verified conditions (1.2) to (1.7) in Proposition I.3. Since γ_0 is the average cost of the "never-turn-on-the-machine" policy, it follows from Proposition I.3 that the said policy is optimal among all policies in $\mathcal{A}_g$. $\square$

Remark I.16. *It will be seen in Section 1.4 that any (s, S) policy belongs to $\mathcal{A}_g$.*

Remark I.17. *In Theorem I.14, the (s^*, S^*) policy with $S^* \geq s^* > 0$ is proved to be optimal when (1.24) and (1.25) are satisfied. If (1.25) is an equality, i.e.,*

$$\int_0^{S_0} (G(x) - \gamma_0 H(x)) dx = -K,$$

then it can be shown that the $(0, S^*)$ *policy and the "never-turn-on-the-machine" policy are both optimal. The proof in Theorem I.14 remains valid.*

Remark I.18. *It can be seen from Figure 1.2(a) that, when* $\sigma_0 = \sigma_1 = 2, c = 7$ *and* $K \geq 5$*, the optimal* s^* *is equal to* 0*. The system parameters in this case satisfy condition (1.24) but not (1.25). Figure 1.2(c) shows that when* $\sigma_0 = 2, \sigma_1 = 3$*,* $c = 7$*, condition (1.24) is not satisfied, i.e.,* $c \not> c_0 = 10$*, therefore in this case the optimal policy satisfies* $s^*(K) = 0$ *for the* K *values considered in this figure. When* $c = 14$*, it would satisfy condition (1.24), but condition (1.25) is still not satisfied for the smallest* K $(K = 1)$ *considered in our numerical test, hence* s^* *is equal to* 0 *on the graph with* $K \geq 1$*. In all these scenarios, "never-turn-on-the-machine" is the optimal policy.*

Theorems I.14 and I.15 state that an (s, S) or the "never-turn-on-the-machine" policy is optimal among the corresponding classes of policies in $\mathcal{A}_v$ or $\mathcal{A}_g$. $\mathcal{A}_v$ and $\mathcal{A}_g$ are dependent on the functions v and g, which is not desired, so we want to know how large the sets of policies $\mathcal{A}_v$ and $\mathcal{A}_g$ are without referring to v and g. In Section 1.4 we present a subset of policies in $\mathcal{A}_v$ that is independent of v, and it contains most of the policies of practical interest.

1.3 The backlog model.

In this section we study the backlog model. Several special cases of the backlogging model have been analyzed in the literature, e.g., Vickson [42] and Doshi [11]. The backlog model is in general simpler to analyze than the lost-sales model, and in this section, we show that regardless of the system parameters, an (s, S) policy is optimal within a large class of policies. The approach we use for the study of the backlog model is similar to that for the lost-sales case, thus most proofs are omitted or put in the appendix.

In the backlog case, the state of the system is still (X_t, Y_t), where X_t is the inventory level and Y_t the mode of production, with $X_t \geq 0$ representing inventory on hand while $X_t < 0$ represents backlog level of $-X_t$. The stochastic process X_t for the inventory level evolves according to the production mode Y_t:

$$\begin{aligned} dX_t &= -\mu_0 dt + \sigma_0 dW_t^0, \quad \text{if } Y_t = 0; \\ dX_t &= \mu_1 dt + \sigma_1 dW_t^1, \quad \text{if } Y_t = 1. \end{aligned}$$

The state space is now $\{(x, y); -\infty < x < \infty, y = 0, 1\}$.

The cost structure is similar to the lost-sales model, except that when $X(t) < 0$, there is a shortage cost rate $h(X_t)$. We make the following assumptions on the holding and shortage cost rate function $h(x)$.

Assumption I.19. *$h(\cdot)$ satisfies*

(i) $h(\cdot)$ is convex;

(ii) $h(\cdot)$ is differentiable;

(iii) $h(\cdot)$ is polynomially bounded; and

(iv) $\lim_{|x|\to\infty} h(x) = +\infty$.

For a policy $\pi \in \mathcal{A}$, with the initial condition (x, y), the expected total cost up to time T is

$$J_{x,y}^{\pi}(T) := E_{x,y}^{\pi}\left[\int_0^T h(X_t)dt + \sum_{0\leq s\leq T} K\delta^+(\Delta Y_s)\right],$$

and the average cost is defined similarly as in the lost-sales model by

$$AC_{x,y}^{\pi} := \limsup_{T\to\infty} \frac{J_{x,y}^{\pi}(T)}{T}.$$

The objective is again to find the optimal policy that minimizes the average cost.

As in §1.2.2, we present two propositions for the backlog model, in parallel to Propositions I.3, I.4. If we can find a function $f(x, y)$ satisfying a set of inequalities, then it yields a lower bound for the long-run average cost.

Proposition I.20. *Suppose that $f(x, y) : \mathbb{R} \times \{0, 1\} \to \mathbb{R}$ is continuously differentiable, and has a continuous second derivative at all but a finite number of points with respect to x. Then for each time $T > 0$, initial state $x \in \mathbb{R}, y \in \{0, 1\}$, and policy π.*

$$E^{\pi}_{x,y}\left[f(X_T, Y_T)\right] = f(x, y) + E^{\pi}_{x,y}\left[\int_0^T \left\{\left(\frac{1}{2}\sigma_0^2 f''(X_t, 0) - \mu_0 f'(X_t, 0)\right)(1 - Y_t) + \right.\right.$$
$$\left.\left. + \left(\frac{1}{2}\sigma_1^2 f''(X_t, 1) + \mu_1 f'(X_t, 1)\right) Y_t\right\} dt\right] + E^{\pi}_{x,y}\left[\sum_{0 \le s \le T} \Delta f(X_s, Y_s)\right], \quad (1.33)$$

where f' and f'' are derivatives with respect to x, $\Delta f(X_s, Y_s) = f(X_s, Y_s) - f(X_s, Y_{s^-})$.

Compared with the lost-sales case, there is no boundary condition for the function $f(x, y)$, thus there are two fewer inequalities for establishing the lower bound.

Proposition I.21. *Suppose that function $f(x, y)$ is polynomially bounded with respect to x and it satisfies all the hypotheses in Proposition I.20, and there exists a positive value γ such that the following conditions are satisfied:*

$$\Gamma_0 f(x, 0) + h(x) - \gamma \ \ge\ 0, \quad (1.34)$$
$$\Gamma_1 f(x, 1) + h(x) - \gamma \ \ge\ 0, \quad (1.35)$$
$$f(x, 1) - f(x, 0) \ \ge\ -K, \quad (1.36)$$
$$f(x, 0) - f(x, 1) \ \ge\ 0, \quad (1.37)$$

then γ is the lower bound of the average cost for all the policies in $\mathcal{A}_f$, i.e.,

$$AC^{\pi} = AC^{\pi}_{x,y} \ge \gamma, \ \forall \pi \in \mathcal{A}_f,$$

in which $\mathcal{A}_f$ *is defined as*

$$\mathcal{A}_f := \left\{ \pi \in \mathcal{A} : \liminf_{T\to\infty} \frac{E^{\pi}_{x,y} f(X_T, Y_T)}{T} \le 0,\ \forall x \in \mathbb{R},\ \forall y \in \{0,1\} \right\}.$$

The (s, S) policy in this section differs from that of the lost-sales case in that s is not necessarily nonnegative.

First, we derive the average cost for an arbitrary (s, S) policy in Proposition I.22, the proof of which is attached in the appendix.

Proposition I.22. *For a given* (s, S) *policy, with* $s < S$*, the average cost of this policy is*

$$c(s,S) \ = \ \frac{\int_s^S G(x)dx + K}{m(S-s)}, \tag{1.38}$$

where

$$\begin{aligned} G(x) \ &= \ \frac{2}{\sigma_0^2}\int_x^{\infty} e^{-\lambda_0(\xi - x)} h(\xi) d\xi + \frac{2}{\sigma_1^2}\int_{-\infty}^{x} e^{-\lambda_1(x-\xi)} h(\xi) d\xi, \\ m \ &= \ \frac{1}{\mu_0} + \frac{1}{\mu_1}. \end{aligned}$$

To find the optimal choice of s and S, we need the following lemma, its proof is given in the appendix.

Lemma I.23. *Suppose the cost function* $h(x)$ *satisfies Assumption I.19, then*

(i) $G(x)$ *is convex, and if* $h(x)$ *is strictly convex, then* $G(x)$ *is also strictly convex;*

(ii) $\lim_{x\to\pm\infty} G(x) = \infty$*;*

(iii) $c(s, S)$ *is strictly convex with respect to* s *and* S*.*

Remark I.24. *Since* $c(s, S)$ *is strictly convex, so the optimal choice of* (s, S) *is unique. The convexity of* $c(x, y)$ *has been established in Zipkin [45] and Zhang [44] for some other context.*

Since $G(x)$ is a convex function converging to infinity as $|x| \to \infty$, it has a minimum say y_0. Clearly, for any $\gamma \geq G(y_0)/m$, there are two points, denoted by $s(\gamma)$ and $S(\gamma)$ respectively, such that $s(\gamma) \leq S(\gamma)$ and $G(s(\gamma)) = G(S(\gamma)) = \gamma m$. The optimal s and S that minimize $c(s, S)$ are determined by the following result.

Theorem I.25. *The optimal choice of s^* and S^*, $-\infty < s \leq S < \infty$, is determined by $s^* = s(\gamma^*)$ and $S^* = S^*(\gamma^*)$ where γ^* is the unique γ satisfying*

$$\int_{s(\gamma)}^{S(\gamma)} \Big(G(x) - m\gamma\Big) dx = -K. \tag{1.39}$$

The illustration of (s^*, S^*) and equation (1.39) are given in Figure 1.3. As can be seen, the area between curve $y = G(x)$ and $y = m\gamma$ is equal to K.

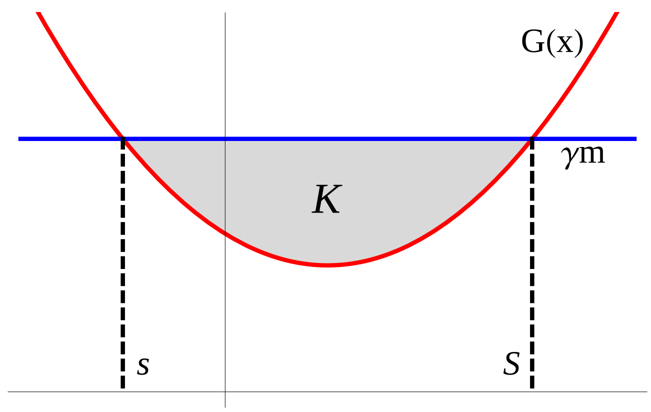


Figure 1.3: Optimal choice of s and S (backlogging case).

With the (s^*, S^*) just identified, we define the relative value functions by

$$v(x,0) = \begin{cases} w_d(x) - \gamma^* \left(\frac{x - s^*}{\mu_0}\right) + v(s^*, 1) + K, & x > s^*; \\ v(x,1) + K, & x \leq s^*, \end{cases} \tag{1.40}$$

$$v(x,1) = \begin{cases} w_u(x) - \gamma^* \left(\frac{S^* - x}{\mu_1}\right), & x < S^*; \\ v(x,0), & x \geq S^*, \end{cases} \tag{1.41}$$

where

$$w_d(x) = \int_{s^*}^{x} \left(\frac{2}{\sigma_0^2} \int_u^{\infty} e^{-\lambda_0(\xi-u)} h(\xi) d\xi \right) du,$$
$$w_u(x) = \int_{x}^{S^*} \left(\frac{2}{\sigma_1^2} \int_{-\infty}^{u} e^{-\lambda_1(u-\xi)} h(\xi) d\xi \right) du$$

are similarly defined as in the lost-sales model. Their continuity can also be similarly shown.

The analysis and optimal control for the backlog model are much simpler than those of the lost-sales model. The main result is that, for backlog model, an (s, S) policy is always optimal within a large class of policies. The approach is similar to the latter part of the lost-sales model; we can show that $v(x, 0)$ and $v(x, 1)$ satisfy all the regularity conditions and inequalities (1.34)-(1.37), thus it follows from γ^* is the average cost for the (s^*, S^*) policy, it must be optimal among the larger class of policies in $\mathcal{A}_v$.

Theorem I.26. *The policy (s^*, S^*) is optimal among the policies in $\mathcal{A}_v$, where v is the relative value function defined in (1.40)-(1.41).*

Theorem I.26 is established under the assumption that the holding and shortage cost rate $h(\cdot)$ is convex. If the production process is deterministic, or $\sigma_0 = \sigma_1$, then we can relax the assumption to quasi-convex[1]. We note that Vickson [42] studies the case of deterministic production process and linear holding cost, and obtains the optimal control policy under certain conditions. The following result extends the results in [42].

Proposition I.27. *If $\sigma_0 = \sigma_1$, then as long as $h(\cdot)$ is quasi-convex, polynomially bounded, and $\lim_{|x|\to\infty} h(x) = \infty$, then an (s, S) policy is optimal among $\mathcal{A}_v$, where v is defined in (1.40)-(1.41).*

[1]A function $f(x)$ is called quasi-convex if for any x, y and $0 \le \lambda \le 1$, $f(\lambda x + (1-\lambda)y) \le \max\{f(x), f(y)\}$.

Proof. From the proceeding analysis, it is not hard to see that all results hold true if $G(x)$ is quasi-convex and $\lim_{|x|\to\infty} G(x) = \infty$. Thus, in the following we show that under the conditions stated in the proposition, $G(x)$ indeed possesses these properties.

As in the proof of Proposition I.22, we obtain

$$\begin{aligned} G(x) &:= \frac{2}{\sigma_0^2}\int_x^\infty e^{-\lambda_0(\xi-x)}h(\xi)d\xi + \frac{2}{\sigma_0^2}\int_{-\infty}^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi \\ &= \frac{2}{\sigma_0^2}\left(\int_0^\infty e^{-\lambda_1 z}h(x-z)dz + \int_{-\infty}^0 e^{\lambda_0 z}h(x-z)dz\right) \\ &= mE\left[h(x-Z)\right], \end{aligned}$$

where Z is a random variable with probability density function

$$p(z) = \begin{cases} \dfrac{1}{m}\dfrac{2}{\sigma_0^2}e^{\lambda_0 z}, & z \le 0; \\ \\ \dfrac{1}{m}\dfrac{2}{\sigma_0^2}e^{-\lambda_1 z}, & z > 0, \end{cases}$$

and, as before, $m = 1/\mu_0 + 1/\mu_1$. Since

$$\log p(z) = \begin{cases} \log\left(2/m\sigma_0^2\right) + \lambda_0 z, & z \le 0; \\ \log\left(2/m\sigma_0^2\right) - \lambda_1 z, & z > 0 \end{cases}$$

is concave, $p(z)$ is log-concave, we deduce that $G(x) = mE[h(x-Z)]$ is quasi-convex [9, p. 17-20] as long as $h(\cdot)$ is quasi-convex. That $\lim_{|x|\to\infty} G(x) = \infty$ is obvious. Thus, the proof of Proposition I.27 is complete. □

Remark I.28. *If $G(x)$ is not strictly quasi-convex*[2]*, then the uniqueness of the optimal (s,S) policy is not guaranteed.*

[2] A function $f(x)$ is strictly quasi-convex if for any x, y such that $f(x) \neq f(y)$ and $0 < \lambda < 1$, $f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$. Alternatively, a function is strictly quasi-convex means that it first strictly decreases and then strictly increases.

1.4 Discussion on $\mathcal{A}_v$ and an example.

The optimality of policies in the previous sections relies heavily on the class of policies $\mathcal{A}_v$. Since v is the relative value function of the said policy, $\mathcal{A}_v$ depends on that policy as well. This does not inform us immediately how large the class of policies $\mathcal{A}_v$ is. We want to know how large this set is without referring to value function v. In this section, we present a subset of $\mathcal{A}_v$ that is independent of v, provided v satisfies some mild conditions. We also discuss scenarios where some policies do not belong to $\mathcal{A}_v$.

Recall that $\mathcal{A}_v$ is defined as the set of admissible policies π for which it holds that

$$\liminf_{T\to\infty} \frac{E^{\pi}_{x,y}\left[f(X_T, Y_T)\right]}{T} \leq 0. \tag{1.42}$$

A similar condition is needed in establishing the optimal policy for production/inventory control problems with infinite capacity and impulse control. The approach used in impulse control is to show that, when this condition is violated by a policy, then that policy has to be a "bad" one. That is, if condition (1.42) is not satisfied by a policy π, then the average cost for policy π is equal to infinity. For example, in Ormeci et al. [30], it is shown that when the holding and shortage cost rate function $h(x)$ is linear, then the relative value function $f(x, y)$ is linearly bounded for $y = 0, 1$. That argument can be extended to the case when $h(x)$ is polynomially bounded of degree n, and it can be shown that the relative value functions for the optimal band policy, $f(x, y)$, are also polynomially bounded functions with the same degree n, and the similar argument as that in Ormeci et al. [30] can be used to show that if a policy does not satisfy (1.42), then the average cost for that policy has to be infinity, see Dai and Yao [23]. As a result, the argument shows that for impulse control problems, the policies not in $\mathcal{A}_v$ can be ignored, implying that a policy that is optimal within the class of policies in $\mathcal{A}_f$ is also optimal among all admissible policies.

One might expect that this argument could be extended to the case with finite production capacity. Unfortunately, that is not the case. When $h(x)$ is polynomially bounded with degree n, the relative value function $f(x, y)$ for an optimal (s, S) policy can be shown to be also polynomially bounded but with degree $n+1$, and violation of (1.42) cannot be used to show that the minimum cost for policy π is infinity. In the following, we first provide an example to demonstrate this, and then we present a subclass of policies that are contained in $\mathcal{A}_v$ but independent of v, and yet it is still large enough to include most policies of practical interest.

Example I.29. Consider a deterministic system: $\mu_0 = \mu_1 = 1$, and $\sigma_1 = \sigma_2 = 0$. The holding cost function is $h(x) = |x|$ and the setup cost is $K > 0$. By choosing the holding cost function in this way, the total holding cost can be interpreted as the area in between X_t and the axis $x = 0$. As for the (s, S) policy, due to the symmetric property of the problem, $s = -S$. As σ_0 and σ_1 converge to 0, we note that $G(x) \to 2h(x)$, thus the average cost for a $(-S, S)$ policy, denoted by $c(S)$, according to (1.38) can be expressed as

$$c(S) = \frac{S}{2} + \frac{K}{4S},$$

the minimum of which is achieved by choosing

$$S = \sqrt{\frac{K}{2}},\ \gamma = \sqrt{\frac{K}{2}}.$$

Let this (s, S) policy be denoted by π. The relative value function for policy π, according to (1.40), (1.41), is

$$v(x,0) = \begin{cases} \frac{x^2}{2} - \sqrt{\frac{K}{2}}x + \frac{K}{4}, & x \geq 0; \\ -\frac{x^2}{2} - \sqrt{\frac{K}{2}}x + \frac{K}{4}, & 0 \geq x \geq -S; \\ v(x,1) + K, & x < -S. \end{cases}$$

$$v(x,1) = \begin{cases} \frac{x^2}{2} + \sqrt{\frac{K}{2}}x - \frac{K}{4}, & x \leq 0; \\ -\frac{x^2}{2} + \sqrt{\frac{K}{2}}x - \frac{K}{4}, & 0 \leq x \leq S; \\ v(x,0), & x > S. \end{cases}$$

Assume ψ denotes the policy of keeping $Y_t = 1$ for all t. If ψ is adopted, then the long-run average cost is infinity. We now construct a policy ϕ using policies π and ψ. Suppose the initial condition is $X_0 = 0$, $Y_{0^-} = 1$, and we construct a policy ϕ as follows. Let $T_n = 2^n$, $n \in \mathbb{Z}^+$. For a sample path ω, if $X_{T_n}(\omega) = T_n$, then the policy for ω is switched to π at time T_n with probability $1/2$, and continue to use ψ until T_{n+1} with probability $1/2$. The evolution of several sample paths are shown in Figure 1.4.

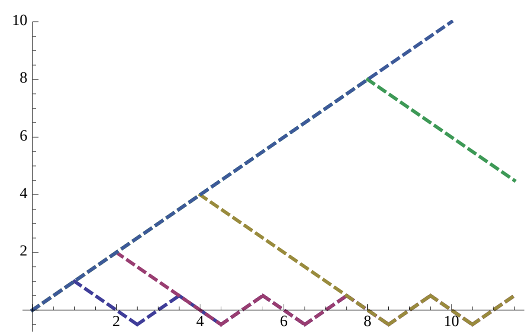


Figure 1.4: Sample paths of ϕ.

For an arbitrary point in time T, there exists an $n \in \mathbb{Z}^+$ such that $2^{n-1} < T \leq 2^n$. Note that $v(x,y) \geq 0$. By considering the top sample path which does not converge to π, we have

$$\begin{aligned}
&\liminf_{T\to\infty} \frac{E\left[v(X_T, Y_T)\right]}{T} \\
\geq\ &\liminf_{T\to\infty} \frac{v(T,1)/2^n}{T} \\
=\ &\liminf_{T\to\infty} \frac{1}{T2^n}\left[\frac{T^2}{2} - \sqrt{\frac{K}{2}}T + \frac{1}{4}K\right] \\
=\ &\liminf_{T\to\infty} \frac{1}{4}\frac{T}{2^{n-1}} \\
\geq\ &\frac{1}{4} \neq 0.
\end{aligned}$$

Thus, this policy ϕ does not belong to $\mathcal{A}_v$. However, by summing up the total expected cost up to time T for all the possible sample paths and using a relaxation, we obtain

$$J_{0,0}^{\phi}(T) \leq \frac{1}{2^n}\left(T^2/2\right) + \frac{1}{2^n}\left[T^2/2 - \left(T - 2^{n-1}\right)^2\right] + \sum_{i=0}^{n-2}\frac{1}{2^{i+1}}\left[(2^i)^2 + \gamma(T - 2\cdot 2^i) + \gamma\cdot 2\sqrt{2K}\right],$$

$$\limsup_{T\to\infty}\frac{J(T)}{T} = \limsup_{T\to\infty}\left\{\frac{T}{2^n} - \frac{(T-2^{n-1})^2}{2^nT} + \frac{2^{n-2}}{T} + \gamma\left(1 - \frac{1}{2^{n-1}}\right)\right\} \leq \gamma + \frac{3}{2}.$$

This shows that there exists non-anticipating policies which do not satisfy the condition $\liminf_{T\to\infty} E[v(X_T, Y_T)]/T = 0$, but it has a finite average cost.

The next question is, how large is the class of policies $\mathcal{A}_v$? We now present a subset of $\mathcal{A}_v$ that is independent of v, and it contains most of the policies of practical interest.

Define

$$\mathcal{A}_\infty := \bigcup_{N>0}\mathcal{A}_N,$$

where $\mathcal{A}_N$ is the class of policies such that $Y_t = 0$ whenever $X_t > N$ and $Y_t = 1$ whenever $X_t < -N$. The definition of $\mathcal{A}_\infty$ does not depend on any specific function; it requires that, as the inventory level becomes very high, the policy should turn off the machine, while when the backlog level becomes very high, then it should turn on the machine. Clearly, most practical policies satisfy this. Further, all (s, S) policies are in $\mathcal{A}_\infty$ too, because under an (s, S) policy, the machine is turned on whenever the inventory level drops to s and turned off whenever the inventory level reaches S.

In the appendix, we show that if $h(\cdot)$ is polynomially bounded with degree n, then $v(x, y)$ is polynomially bounded with degree $n + 1$. Under the condition that $v(x, y)$ is polynomially bounded, we will prove $\mathcal{A}_\infty \subset \mathcal{A}_v$.

Proposition I.30. *$\mathcal{A}_\infty \subset \mathcal{A}_v$ if v is polynomially bounded, i.e., there exists an $n \in \mathbb{Z}^+$ such that $|v(x, y)| \leq \bar{v}(x) := B_1 + B_2|x|^n$, for some constants B_1 and B_2 and $\forall x \in \mathbb{R}$, $\forall y \in \{0, 1\}$.*

Proof. For a given policy $\pi \in \mathcal{A}_\infty$, there exists an N such that $\pi \in \mathcal{A}_N$. Without loss of generality, suppose the initial condition $X_0 = x \in (-N, N)$. We construct a process $M_1(t)$ as follows

$$M_1(t) = X_t, \text{ for } 0 \leq t \leq \tau_N,$$

where τ_N is the hitting time of N; and after hitting time τ_N, $M_1(t)$ is a Brownian motion with downward drift $-\mu_0$, variance σ_0^2, and N is the one-sided reflecting lower boundary, that is the process $M_1(t)$ is always at or above N after time τ_N. It is easy to show that $X_t \leq M_1(t)$. Similarly, we construct another process $M_2(t)$ by

$$M_2(t) = X_t, \text{ for } 0 \leq t \leq \tau_{-N},$$

where τ_{-N} is the hitting time of $-N$; and after hitting time τ_{-N}, let $M_2(t)$ be a Brownian motion with upward drift μ_1, variance σ_1^2, and with $-N$ as a one-sided reflecting upper boundary, thus $M_2(t)$ will be always at or below $-N$ after τ_{-N}. It can be seen that $X_t \geq M_2(t)$. If $X_t \geq 0$, then $\bar{v}(X_t) \leq \bar{v}(M_1(t))$; and if $X_t < 0$, then

$\bar{v}(X_t) \le \bar{v}(M_2(t))$. Thus, $\bar{v}(X_t) \le \max\{\bar{v}(M_1(t)), \bar{v}(M_2(t))$ for all t, and

$$\begin{aligned}\frac{E^{\pi}_{x,y}\left[v(X_T, Y_T)\right]}{T} &\le \frac{E^{\pi}_{x,y}\left[\bar{v}(X_T)\right]}{T} \\ &\le \frac{E^{\pi}_{x,y}\left[\max\left\{\bar{v}(M_1(T)), \bar{v}(M_2(T))\right\}\right]}{T} \\ &\le \frac{E^{\pi}_{x,y}\left[\bar{v}(M_1(T))\right] + E^{\pi}_{x,y}\left[\bar{v}(M_2(T))\right]}{T}.\end{aligned}$$

Regulated Brownian motion processes with one-side reflecting boundary are well understood, see e.g., Harrison [19, §5.6]. Since both $M_1(t)$ and $M_2(t)$ have exponential steady state distributions, that is, the probability for $M_1(t)$ ($M_2(t)$) to take large (negative) value is exponentially decaying. This shows that the numerator of the fraction above is finite, so after taking lower limit we obtain, when $\bar{v}(x)$ is polynomially bounded,

$$\liminf_{T\to\infty} \frac{E^{\pi}_{x,y}\left[v(X_T, Y_T)\right]}{T} \le 0.$$

This proves that $\mathcal{A}_\infty \subset \mathcal{A}_v$ for all v polynomially bounded. $\square$

Remark I.31. *The definition of $\mathcal{A}_\infty$ above is similar to that defined in [41].*

Remark I.32. *The argument above can be used to show that, actually,*

$$\lim_{T\to\infty} \frac{E^{\pi}_{x,y}\left[v(X_T, Y_T)\right]}{T} = 0 \text{ for all } \pi \in \mathcal{A}_\infty.$$

Remark I.33. *The parameters for the exponential steady state distributions of $M_1(t)$ and $-M_2(t)$ are $\sigma_0^2/(2\mu_0)$ and $\sigma_1^2/(2\mu_1)$, respectively. It follows from the argument above that the result would be true as long as $\bar{v}(x)$ is bounded by an exponential function with parameter less than $\min\{\sigma_0^2/(2\mu_0), \sigma_1^2/(2\mu_1)\}$.*

Remark I.34. *For the lost-sales model, the subset of policies can be defined as $\mathcal{A}_\infty = \cup_{N>0}\mathcal{A}_N$, where $\mathcal{A}_N$ contains all the policies satisfying that whenever $X_t \ge N$, $Y_t = 0$. The same argument used above shows that $\mathcal{A}_\infty \subset \mathcal{A}_v$ for all v polynomially bounded.*

CHAPTER II

Bound on the Coarsening Rate and Classical Limit Theorem for the Diffusive Carr-Penrose Model

2.1 Introduction

Ostwald ripening (Coarsening) is a physics phenomenon observed in solid solutions or liquid sols. Since monomers (single particles) have a larger surface area, thus energetically less stable compared with polymers (clusters of particles), they tend to be absorbed by polymers. Similarly, polymers with a small amount of particles tend to have their surface particles detached from them; polymers with a large amount of particles are formed thus to achieve higher stability. This process of smaller polymers shrinking, while larger polymers growing, with the average size of the system increasing is called Ostwald ripening, which was first described by Wilhelm Ostwald in 1896.

Ostwald ripening is an important phenomenon since it occurs in crystallization, coarsening of sorted stone stripes, synthesis of quantum dots, coalescence of alloy, supersaturated solutions, digestion of precipitates, emulsion systems etc. Lifshitz, Slyozov and Wagner are pioneers in this field of research. In 1961, Lifshitz and Slyozov jointly and independently Wagner developed theories to explain the phenomenon. Their conclusions (though obtained using different methods) were shown to be the same by Kahlweit in 1975, and are referred to as the Lifshitz-Slyozov-Wagner (LSW) Theory of Ostwald ripening. The main focus of it is on the description of the density (or concentration) function of polymers of different sizes at large time as well as the coarsening rate-the rate at which the average size increases.

LSW theory involves solving a *nonlinear nonlocal* first order partial differential equation (PDE) which in general does not have explicit solutions. Though self-similar solutions are identified and predicted as long time asymptotes of a general initial condition, the intractability of the nonlinear differential equation itself still hinders a clear understanding of the solution. Carr and Penrose (1998) [3] propose a linear version of the PDE, which is tractable. In their paper, they show that for a large class of initial data, the solution behaves asymptotically like one of the self-similar solutions, and which solution it converges to depends solely on the behavior of the initial condition towards the end of its support. The same conclusion is believed to be true for the LSW equation. Meerson (1999) [28] argues that by adding a diffusive term to the LSW PDE, which adds a "Gaussian tail" to the initial condition, *a strong selection principle* is obtained. When applied to the Carr-Penrose (CP) model, a similar result should hold, i.e., only the exponential self-similar solution should give the asymptotic behavior for the solutions of the CP model.

In this chapter, we study a CP model with a diffusive term. We express the solution to the diffusive CP partial differential equation using a Dirichlet Green's function, and present the connection between the Dirichlet Green's function and the characteristic solution to the classical CP model. Then, we link the Dirichlet Green's function with the distribution function of a Gaussian process which has fixed initial and terminal conditions. Instead of using the Markovian representation of this process, which works well for constant drift cases, we adopt a non-Markovian representation. We use it to show the convergence of the density function of the diffusive CP model to the classical one as the diffusion constant $\varepsilon \to 0$. In order to show the convergence of the coarsening rate, we derive uniform (in terms of the diffusion constant $\varepsilon > 0$) bounds (upper and lower) of the ratio between the Dirichlet Green's function to the full space Green's function. Due to the non-Markovian nature of the representation of the Gaussian process, the value of the process at a certain time point depends on both the realization of a Brownian motion in the past and the future. In the derivation of the bounds, we use two main techniques (observations): the considered stochastic process should be compared with a tractable approximating process which has a constant drift; based on different realization of the Brownian motion part of the stochastic process, the drift of the process to compare with should vary. (See Lemma II.18, II.20, II.22.) Last, we demonstrate the connection between log concavity of the initial condition and a *beta function* first defined in Conlon (2011) [7], and the relation between the coarsening rate and this beta function. With a log concavity assumption on the initial condition, we derive an upper bound on the coarsening rate by using this beta function and the bounds on the ratio between the Dirichlet and the full space Green's function.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the Carr-Penrose model and its explicit solution. In Section 2.3, we introduce the diffusive CP model, study its general solution and estimate the Dirichlet Green's function. Then in Section 2.4, the convergence of the solution and coarsening rate of the diffusive CP model to the classical case is studied. Finally, we derive an upper bound on the coarsening rate for the diffusive CP model given certain log concavity for the initial condition in Section 2.5.

2.2 Classical Carr-Penrose model

2.2.1 The problem

In the theory of coarsening, the system of differential equations to characterize the concentration of the clusters of different sizes can be expressed as in [3]:

$$\frac{\partial c(x,t)}{\partial t} = -\frac{\partial}{\partial x}[v(x,t)c(x,t)], \tag{2.1}$$

$$v(x,t) = a(x)[1/\Lambda(t) - x^{-1/\nu}], \tag{2.2}$$

$$\int_0^\infty xc(x,t)dx = 1, \tag{2.3}$$

where $x, t \geq 0$ and $c(x,t)$ represents the density (concentration), at time t, of clusters consisting of x particles, $v(x,t)$ is the average rate at which the number of particles in a cluster grows, $a(x)$ is a given function of x, and ν is the number of space dimensions. The mass conservation equation (2.3) implies that

$$\Lambda(t) = \frac{\int_0^\infty a(x)c(x,t)dx}{\int_0^\infty x^{-1/\nu}a(x)c(x,t)dx}.$$

We assume that $a(x)$ is proportional to a power of x, say $a(x) = \alpha x^{\beta+1/\nu}$ where α and β are positive constants.

If we choose $\alpha = 1$, then equations (2.1), (2.2) together with (2.2.1) give

$$\begin{aligned} \frac{\partial c(x,t)}{\partial t} &= -\frac{\partial}{\partial x}\{x^\beta[1/\Lambda(t)\cdot x^{1/\nu} - 1]c(x,t)\}, \\ \Lambda(t) &= \frac{\int_0^\infty x^{\beta+1/\nu}c(x,t)dx}{\int_0^\infty x^\beta c(x,t)dx} \end{aligned}$$

with the conservation law

$$\int_0^\infty xc(x,t)dx = 1. \tag{2.4}$$

If we choose $\beta = 0$, $\nu = 3$, then the system is the Lifshitz-Slyozov-Wagner (LSW) model; if $\beta = 0, \nu = 1$, it becomes the Carr-Penrose model [3], which we will discuss below. For Carr-Penrose model, we have

$$\frac{\partial c}{\partial t} = -\frac{\partial}{\partial x}\left\{\left[\frac{x}{\Lambda(t)} - 1\right] c(x,t)\right\}, \tag{2.5}$$

$$\Lambda(t) = \frac{\int_0^\infty x \cdot c(x,t)dx}{\int_0^\infty c(x,t)dx}. \tag{2.6}$$

with (2.4) $\int_0^\infty xc(x,t)dx = 1$. We can define a random variable X_t whose probability density function is given by $c(x,t)/\int_0^\infty c(x,t)dx$. Then the mean of X_t, $\langle X_t \rangle = \Lambda(t)$[1], in which $\Lambda(t)$ is a continuous function.

2.2.2 General solution

We define the function $w(x,t) = \int_x^\infty c(x',t)dx'$. It is easy to see that

$$\frac{w(x,t)}{w(0,t)} = \frac{\int_x^\infty c(x',t)dx'}{\int_0^\infty c(x,t)dx} = P(X_t \geq x). \tag{2.7}$$

Following (2.4), the conservation law for $w(x,t)$ becomes

$$\int_0^\infty w(x,t)dx = 1. \tag{2.8}$$

Also, we define a function $h(x,t) = \int_x^\infty w(x',t)dx'$. Due to conservation law (2.4), $\langle X_t \rangle = 1/w(0,t)$,

$$h(x,t) = \frac{\int_x^\infty w(x',t)dx'}{w(0,t)} \cdot w(0,t) = \frac{\int_x^\infty P(X_t > x')dx'}{\langle X_t \rangle} = \frac{E[X_t - x; X_t > x]}{\langle X_t \rangle},$$

and it follows from the conservation law (2.8) that

$$h(0,t) = 1. \tag{2.9}$$

[1] $\langle \cdot \rangle$ denotes the mean of a random variable.

Due to the differential equation (2.5) for $c(x,t)$, we have a corresponding differential equation for $w(x,t)$:

$$\frac{\partial w}{\partial t} + \left(\frac{x}{\Lambda(t)} - 1\right)\frac{\partial w}{\partial x} = 0, \quad w(x,0) = w_0(x) := \int_x^\infty c_0(x)dx. \tag{2.10}$$

Lemma II.1. *If $\Lambda(t)$ is continuous, then the general solution of $w(x,t)$ to (2.10) is*

$$w(x,t) = w_0(F(x,t)), \tag{2.11}$$

where

$$F(x,t) = u(t)x + v(t) \tag{2.12}$$

and $u(t), v(t)$ are defined by

$$u(t) = \exp\left[-\int_0^t \frac{1}{\Lambda(s)}ds\right],\ v(t) = \int_0^t \exp\left[-\int_0^s \frac{1}{\Lambda(s')}ds'\right]ds. \tag{2.13}$$

Proof. We use the method of characteristic, and consider a curve $x(s)$ satisfying the condition

$$\frac{d}{dt}x(t) = \frac{x(t)}{\Lambda(t)} - 1, \quad x(0) = x_0.$$

The solution to this initial value ordinary differential equation problem is

$$x(t) = x_0 \cdot \exp\left[\int_0^t \frac{1}{\Lambda(s)}ds\right] - \int_0^t \exp\left[\int_{s'}^t \frac{1}{\Lambda(s'')}ds''\right]ds',$$

or, in another form,

$$x_0 = \exp\left[-\int_0^t \frac{1}{\Lambda(s)}ds\right]x(t) + \int_0^t \exp\left[-\int_0^s \frac{1}{\Lambda(s')}ds'\right]ds = F(x(t),t).$$

Since over this characteristic curve, we have $\frac{d}{dt}w(x(t),t) = 0$, it follows that

$$w(x(t),t) = w(x_0, 0) = w_0(F(x(t),t)),$$

thus the lemma is proved. $\square$

It follows from the relation between $w(x,t)$ and $c(x,t)$ that

$$c(x,t) = c_0(F(x,t)) \cdot u(t). \tag{2.14}$$

Similarly,

$$h(x,t) = \frac{1}{u(t)} h_0(F(x,t)), \tag{2.15}$$

where $h_0(\cdot) := h(\cdot, 0)$.

In [6], Conlon shows the existence of the solution to a diffusive Lifschitz-Slyozov-Wagner equation. Actually, for the classical Carrr-Penrose model, a simpler case, we can associate the existence of $\Lambda(t)$, thus $c(x,t)$, with a 2-dim dynamics.

Lemma II.2. *For Carr-Penrose model* (2.5) *with conservation law*

$$\int_0^\infty x \cdot c(x,t) dx = 1,$$

if $c_0(\cdot)$ is bounded, then $\Lambda(t)$ and $u(t), v(t)$ defined in (2.13) *exist.*

Proof. Consider an equation (due to the conservation law),

$$\int_0^\infty w_0(u(t)x + v(t)) dx = 1.$$

Taking the derivative and we have

$$\frac{d}{dt}\left[\int_0^\infty w_0(u(t)x + v(t)) dx\right] = 0,$$

i.e.,

$$\frac{du}{dt}\int_0^\infty x\ w_0'(u(t)x+v(t))dx+\frac{dv}{dt}\int_0^\infty w_0'(u(t)x+v(t))dx=0.$$

By substitution $z=u(t)x$,

$$\frac{1}{u^2(t)}\frac{du}{dt}\int_0^\infty z\cdot w_0'(z+v(t))dz+\frac{1}{u(t)}\cdot\frac{dv}{dt}\int_0^\infty w_0'(z+v(t))dz=0.$$

By integration by parts,

$$-\frac{1}{u^2(t)}\frac{du}{dt}\int_0^\infty w_0(z+v(t))dz-\frac{1}{u(t)}\frac{dv}{dt}w_0(v(t))=0.$$

Therefore, we have a system of ordinary differential equations

$$\frac{dv}{dt}=u(t),\frac{du}{dt}=-\frac{w_0(v(t))u^2(t)}{\int_0^\infty w_0(z+v(t))dz},\quad u(0)=1,v(0)=0. \tag{2.16}$$

As long as w_0 is Lipschitz continuous, the ordinary differential equation system has a unique solution. The Lipschitz continuity can be guaranteed by the boundedness of c_0, under which condition, the solutions for u and v exist and thus $\Lambda(t)$ exists. Also, $\Lambda(t)$ is continuous due to the continuity of $u(t)$ and $v(t)$. □

2.2.3 Coarsening rate

We define

$$\beta(x,t)=\frac{c(x,t)h(x,t)}{w(x,t)^2}, \tag{2.17}$$

as in [7] and it follows from (2.14) and (2.15) that

$$\beta(x,t)=\beta(F(x,t),0). \tag{2.18}$$

The definition of $\beta(x,t)$ connects to the coarsening rate, which is shown by the following lemma.

Lemma II.3. *The mean of X_t, $\Lambda(t)$, has a derivative $\beta(0,t)$, i.e.,*

$$\frac{d\Lambda(t)}{dt} = \beta(0,t). \tag{2.19}$$

Proof.

$$\frac{d\Lambda(t)}{dt} = \frac{d}{dt}\left[\frac{1}{w(0,t)}\right] = \frac{-\frac{\partial w}{\partial t}(0,t)}{w(0,t)^2} \tag{2.20}$$

Since $w(x,t)$ satisfies the partial differential equation (2.10), $\frac{\partial w}{\partial t}(0,t) = \frac{\partial w}{\partial x}(0,t)$. Therefore, by recalling the conservation law for $h(x,t)$, $h(0,t) = 1$, from (2.9),

$$\frac{d\Lambda(t)}{dt} = \frac{-1 \cdot \frac{\partial w}{\partial x}(0,t)}{w(0,t)^2} = \frac{h(0,t)c(0,t)}{w(0,t)^2} = \beta(0,t).$$

□

Due to (2.17), we expect the coarsening rate to be determined by the behavior of $\beta(x_\infty, 0)$ with $x_\infty := \sup_x\{w_0(x) > 0\}$.

Lemma II.4. *Suppose $x_\infty = \sup_x\{w_0(x) > 0\}$, then*

$$\lim_{t\to\infty} F(0,t) = x_\infty. \tag{2.21}$$

Proof. We know

$$\Lambda(t) = \langle X_t \rangle = \frac{1}{w(0,t)}.$$

Since $F(0,t)$ is an increasing function (see (2.13)), $\lim_{t\to\infty} F(0,t)$ always exists ($+\infty$ is included). If $\lim_{t\to\infty} F(0,t) = x_1 < x_\infty$, then since $w(0,t) \geq w(x_1, 0) > 0$ for all t,

$$\Lambda(t) \leq \Lambda_\infty := \frac{1}{w(x_1,0)}, \text{ for all } t.$$

Therefore, $\lim_{t\to\infty} u(t) = 0$. According to the conservation law,

$$\int_0^\infty w(x,t)dx = \int_0^\infty w_0(F(x,t))dx = 1,$$

which simplifies as

$$\int_0^\infty w_0(u(t)x + v(t), 0)dx = 1.$$

By substitution $z = u(t)x + v(t)$,

$$\int_{v(t)}^\infty w_0(z)dz = u(t).$$

Let $t \to \infty$, the left hand side approaches $\int_{x_1}^\infty w_0(z)dz > 0$ and the right hand side approaches 0, leading to a contradiction. Hence, $\lim_{t\to\infty} F(0,t) \geq x_\infty$.

On the other side, we notice that the characteristic curves are level sets for the function $w(x,t)$. If $\lim_{t\to\infty} F(0,t) > x_\infty$, then there exists t^* such that $F(0,t^*) = x_\infty$, implying that $w(0,t^*) = w(x_\infty, 0) = 0$, a contradiction to $w(0,t^*) > 0$. Therefore $\lim_{t\to\infty} F(0,t) \leq x_\infty$. It follows that $\lim_{t\to\infty} F(0,t) = x_\infty$. $\square$

In the following, we use β_0 to denote the limit of $\beta(x,0)$ as x approaches the upper limit of the support of $w_0(\cdot)$: $\beta_0 := \lim_{x\to x_\infty} \beta(x,0)$. Therefore, $\lim_{t\to\infty} \beta(0,t) = \beta_0$, and the asymptotic behavior of the coarsening rate $d\Lambda(t)/dt$ is determined by β_0. In the following, we give three examples demonstrating initial conditions with different β_0 values.

Example II.5. For the function $w_0(x) = (\alpha+1)(1-x)^\alpha$ with $\alpha > 0$, $x_\infty = 1$. We have

$$c_0(x) = \alpha(\alpha+1)(1-x)^{\alpha-1}, \quad h_0(x) = (1-x)^{\alpha+1}.$$

Therefore, the beta function at $t = 0$ is

$$\beta(x,0) = \frac{\alpha(\alpha+1)(1-x)^{\alpha-1} \cdot (1-x)^{\alpha+1}}{(\alpha+1)^2(1-x)^{2\alpha}} = \frac{\alpha}{\alpha+1}.$$

Thus, $\beta_0 = \alpha/(\alpha+1) < 1$.

Example II.6. For $w_0(x) = (\alpha-1)/(x+1)^\alpha$ with $\alpha > 1$, $x_\infty = \infty$.

$$c_0(x) = \frac{\alpha(\alpha-1)}{(1+x)^{\alpha+1}}, \quad h_0(x) = \frac{1}{(1+x)^{\alpha-1}}.$$

Therefore,

$$\beta(x,0) = \frac{\frac{\alpha(\alpha-1)}{(1+x)^{\alpha+1}} \cdot \frac{1}{(1+x)^{\alpha-1}}}{(\alpha-1)^2/(x+1)^{2\alpha}} = \frac{\alpha}{\alpha-1}.$$

Thus, $\beta_0 = \alpha/(\alpha-1) > 1$.

Example II.7. For $w_0(x) = e^{-x}$, $x_\infty = \infty$.

$$c_0(x) = e^{-x}, \quad h_0(x) = e^{-x}.$$

Therefore, $\beta(x,0) = 1$ and $\beta_0 = 1$.

2.3 Diffusive CP model

2.3.1 The problem

In this section, we study a diffusive version of the Carr-Penrose model. Let $\varepsilon > 0$ be the diffusion constant, and the density function $c_\varepsilon(x,t)$ satisfies the differential equation together with the conservation constraint as follows:

$$\frac{\partial c_\varepsilon(x,t)}{\partial t} + \frac{\partial}{\partial x}\left[\frac{x}{\Lambda_\varepsilon(t)} - 1\right] c_\varepsilon(x,t) = \frac{\varepsilon}{2}\frac{\partial^2 c_\varepsilon}{\partial x^2}. \tag{2.22}$$

$$\int_0^\infty x c_\varepsilon(x,t)dx = 1, \quad c_\varepsilon(0,t) = 0. \tag{2.23}$$

Similar to the classical CP model, we define $w_\varepsilon(x,t) = \int_x^\infty c_\varepsilon(x',t)dx'$ and $h_\varepsilon(x,t) = \int_x^\infty w(x',t)dx'$. Since the initial condition does not depend on ε, the corresponding initial conditions are still named as $c_0(\cdot), w_0(\cdot)$ and $h_0(\cdot)$. The differential equation $w_\varepsilon(x,t)$ satisfies is

$$\frac{\partial w_\varepsilon}{\partial t} + \left[\frac{x}{\Lambda_\varepsilon(t)} - 1\right]\frac{\partial w_\varepsilon}{\partial x} = \frac{\varepsilon}{2}\frac{\partial^2 w_\varepsilon}{\partial x^2}. \tag{2.24}$$

Identical to the classical CP model, we have

$$\Lambda_\varepsilon(t) = \frac{\int_0^\infty x c_\varepsilon(x,t)dx}{\int_0^\infty c_\varepsilon(x,t)dx}. \tag{2.25}$$

The derivation of (2.25) relies on the Dirichlet condition $c_\varepsilon(0,t) = 0$. Subsequently,

$$\frac{d\Lambda_\varepsilon(t)}{dt} = \frac{-\frac{\partial}{\partial t}w(0,t)}{[\int_0^\infty c_\varepsilon(x,t)dx]^2}.$$

Due to (2.24),

$$\frac{\partial}{\partial t}w(0,t) = \frac{\partial w_\varepsilon(0,t)}{\partial x} + \frac{\varepsilon}{2}\frac{\partial^2 w_\varepsilon(0,t)}{\partial x^2} = -c_\varepsilon(0,t) + \frac{\varepsilon}{2}\frac{\partial^2 w_\varepsilon(0,t))}{\partial x^2} = -\frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,t)}{\partial x},$$

where the last equation follows from the boundary condition $c_\varepsilon(0,t) = 0$. Therefore, the coarsening rate for the diffusive Carr-Penrose model is given by

$$\frac{d\Lambda_\varepsilon(t)}{dt} = \frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,t)}{\partial x} \Bigg/ \left[\int_0^\infty c_\varepsilon(x,t)dx\right]^2. \tag{2.26}$$

Remark II.8. *The difference between the coarsening rate of the diffusive case and the one of the classical case is between $\frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,t)}{\partial x}$ and $c(0,t)$. This will be discussed later in Lemma II.24.*

2.3.2 Representation using the Green's Functions

In the following, we introduce some common theories of Green's function, which helps in the expression of the solution the diffusive CP model.

Let $b : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be a continuous function which satisfies the uniform Lipschitz condition

$$\sup\{|\partial b(y,t)/\partial y| : y, t \in \mathbf{R}\} \le A \tag{2.27}$$

for some constant A. Then the terminal value problem

$$\frac{\partial u_\varepsilon(y,t)}{\partial t} + b(y,t)\frac{\partial u_\varepsilon(y,t)}{\partial y} + \frac{\varepsilon}{2}\frac{\partial^2 u_\varepsilon(y,t)}{\partial y^2} = 0, \quad y \in \mathbf{R}, t < T, \tag{2.28}$$

$$u_\varepsilon(y,T) = u_T(y), y \in \mathbf{R}, \tag{2.29}$$

has a unique solution u_ε which has the representation

$$u_\varepsilon(y,t) = \int_{-\infty}^{\infty} G_\varepsilon(x,y,t,T) u_T(x) dx, \quad y \in \mathbf{R}, t < T, \tag{2.30}$$

where G_ε is the Green's function for the problem. For any $t < T$, let $Y_\varepsilon(s)$, $s > t$, be the solution to the initial value problem for the stochastic differential equation

$$dY_\varepsilon(s) = b(Y_\varepsilon(s), s)ds + \sqrt{\varepsilon} dB(s), \quad Y_\varepsilon(t) = y, \tag{2.31}$$

where $B(\cdot)$ is Brownian motion. Then $G_\varepsilon(\cdot, y, t, T)$ is the probability density for $Y_\varepsilon(T)$. The solution to (2.28) has an expectation expression

$$u_\varepsilon(y,t) = E\left[u_T(Y_\varepsilon(T)) \middle| Y_\varepsilon(t) = y\right]. \tag{2.32}$$

The adjoint problem to (2.28), (2.29) is the initial value problem

$$\frac{\partial v_\varepsilon(x,t)}{\partial t} + \frac{\partial}{\partial x}[b(x,t)v_\varepsilon(x,t)] = \frac{\varepsilon}{2}\frac{\partial^2 v_\varepsilon(x,t)}{\partial x^2}, \quad x \in \mathbf{R}, t > 0, \tag{2.33}$$

$$v_\varepsilon(x,0) = v_0(x), \quad y \in \mathbf{R}. \tag{2.34}$$

The solution to (2.33), (2.34) is given by the formula

$$v_\varepsilon(x,t) = \int_{-\infty}^{\infty} G_\varepsilon(x,y,0,t)v_0(y)dy, \quad x \in \mathbf{R}, t > 0. \tag{2.35}$$

Parallel to (2.31), we consider a diffusion process $X_\varepsilon(\cdot)$ run backwards in time,

$$dX_\varepsilon(s) = b(X_\varepsilon(s),s)ds + \sqrt{\varepsilon}dB(s),\ X_\varepsilon(T) = x,\ s < T, \tag{2.36}$$

where $B(s)$, $s < T$ is Brownian motion run *backwards*. The solution v_ε of (2.33) has an expectation representation

$$v_\varepsilon(x,T) = E\left[\exp\left\{-\int_0^T \frac{\partial b(X_\varepsilon(s),s)}{\partial x}ds\right\} v_0(X_\varepsilon(0))\middle| X_\varepsilon(T) = x\right]. \tag{2.37}$$

Remark II.9. *Here the Green's function satisfies both Kolmogrov backward and forward equations* (2.28)*,* (2.33)*. Below, we explain this further. Let $\mathcal{L}$ be a differential operator defined as*

$$\mathcal{L}u(y,t) := -b(y,t)\frac{\partial}{\partial y}u - \frac{\varepsilon}{2}\frac{\partial^2}{\partial y^2}u.$$

Then (2.28) *can be expressed as $\frac{\partial u}{\partial t} = \mathcal{L}u$. Since*

$$\int_{-\infty}^{\infty} -b(y,t)\frac{\partial u}{\partial y}v(y,t)dy = -b(y,t)u\cdot v\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} u\frac{\partial}{\partial y}(b(y,t)v)dy = \int_{-\infty}^{\infty} u\frac{\partial}{\partial y}(b(y,t)v)dy,$$

and

$$\int_{-\infty}^{\infty} -\frac{\varepsilon}{2}\frac{\partial^2 u}{\partial y^2} v(y,t)dy = -\frac{\varepsilon}{2}\frac{\partial u}{\partial y} v(y,t)\Big|_{-\infty}^{\infty} + \frac{\varepsilon}{2}\int_{-\infty}^{\infty} \frac{\partial u}{\partial y}\cdot\frac{\partial v}{\partial y}dy$$
$$= \frac{\varepsilon}{2} u \cdot \frac{\partial v}{\partial y}\Big|_{-\infty}^{\infty} - \frac{\varepsilon}{2}\int_{-\infty}^{\infty} u\cdot\frac{\partial^2 v}{\partial y^2}dy = -\frac{\varepsilon}{2}\int_{-\infty}^{\infty} u\cdot\frac{\partial^2 v}{\partial y^2}dy,$$

the adjoint of $\mathcal{L}$, which we denote by $\mathcal{L}^*$, is defined as

$$\mathcal{L}^* v(y,t) = \frac{\partial}{\partial y}(b(y,t)v) - \frac{\varepsilon}{2}\frac{\partial^2 v}{\partial y^2},$$

and the adjoint problem of (2.28) is $\frac{\partial v}{\partial t} = -\mathcal{L}^* v$, which is (2.33).

Since $u(x,t)$ and $v(x,t)$ are solutions to adjoint processes,

$$\frac{d}{dt}[u(x,t), v(x,t)] = 0,$$

where $[u,v] := \int_{-\infty}^{\infty} u(x,t)v(x,t)dx$. This implies $[u(x,0), v(x,0)] = [u(x,T), v(x,T)]$. By choosing terminal and initial conditions $u(x,T) = \delta(x-x_0)$ and $v(y,0) = \delta(y-y_0)$, we obtain $u(y_0,0) = v(x_0,T)$. Due to (2.30), $u(y_0,0) = G(x_0,y_0,0,T)$, thus $v(x_0,T) = G(x_0,y_0,0,T)$.

Next, in the case when $b(y,t)$ is linear in y, e.g., $b(y,t) = A(t)y - 1$, where $A : \mathbf{R} \to \mathbf{R}$, the solution to (2.31) is given by

$$Y_\varepsilon(s) = m_1(t,s)y - m_2(t,s) + \sqrt{\varepsilon}\int_t^s \exp\left[\int_{s'}^s A(s'')ds''\right] dB(s'), \tag{2.38}$$

where

$$m_1(t,s) = \exp\left[\int_t^s A(s')ds'\right],\ m_2(t,s) = \int_t^s \exp\left[\int_{s'}^s A(s'')ds''\right] ds'. \tag{2.39}$$

Hence $Y_\varepsilon(s)$ conditioned on $Y_\varepsilon(t) = y$ is Gaussian with mean $m_1(t,s)y - m_2(t,s)$ and variance $\varepsilon\sigma^2(t,s)$ where

$$\sigma^2(t,s) = \int_t^s \exp\left[2\int_{s'}^s A(s'')ds''\right] ds'. \tag{2.40}$$

For future convenience, we also define $m_1(T) = m_1(0,T), m_2(T) = m_2(0,T)$ and $\sigma^2(T) = \sigma^2(0,T)$. Thus the Green's function is expressed as

$$G_\varepsilon(x,y,0,T) = \frac{1}{\sqrt{2\pi\varepsilon\sigma^2(T)}} \exp\left[-\frac{\{x+m_2(T)-m_1(T)y\}^2}{2\varepsilon\sigma^2(T)}\right]. \tag{2.41}$$

Remark II.10. *We note that*

$$\sigma^2(t,s) = \sigma^2(t,t')m_1^2(t',s) + \sigma^2(t',s), \ \textit{for } t < t' < s. \tag{2.42}$$

$$m_2(t,s) = m_2(t,t')m_1(t',s) + m_2(t',s), \ \textit{for } t < t' < s. \tag{2.43}$$

It is convenient to use these relations in derivations of some subsequent results.

Next, we consider the problem (2.28), (2.29) in the half space $y > 0$ with Dirichlet boundary condition $u_\varepsilon(0,t) = 0$, $t < T$. In this case, similar to (2.30), $u_\varepsilon(y,t)$ has the representation

$$u_\varepsilon(y,t) = \int_0^\infty G_{\varepsilon,D}(x,y,t,T)u_T(x)dx, \quad y > 0, t < T, \tag{2.44}$$

where $G_{\varepsilon,D}$ is the Dirichlet Green's function. Similarly, the solution to (2.33), (2.34) in the half space $x > 0$ with Dirichlet condition $v_\varepsilon(0,t) = 0$, $t > 0$ has the representation

$$v_\varepsilon(x,T) = \int_0^\infty G_{\varepsilon,D}(x,y,0,T)v_0(y)dy, \quad x > 0, T > 0. \tag{2.45}$$

In terms of probability, $G_{\varepsilon,D}(\cdot,y,t,T)$ is the probability density function of the random variable $Y_\varepsilon(T)$ satisfying (2.31) conditioned on $\inf_{t\le s\le T} Y_\varepsilon(s) > 0$. In most scenarios, there is no explicit formula for $G_{\varepsilon,D}(x,y,0,T)$. Exceptionally however, when $A(\cdot) \equiv 0$, $G_{\varepsilon,D}(x,y,0,T)$ has an explicit formula by the method of images.

$$G_{\varepsilon,D}(x,y,0,T) = \frac{1}{\sqrt{2\pi\varepsilon T}} \left\{ \exp\left[-\frac{(x-y+T)^2}{2\varepsilon T} \right] - \exp\left[-\frac{2x}{\varepsilon} - \frac{(x+y-T)^2}{2\varepsilon T} \right] \right\} \tag{2.46}$$

Remark II.11. *We note that $G_{\varepsilon,D}(x,y,0,T)$ satisfies the differential equation (2.28), boundary condition $G_{\varepsilon,D}(0,y,0,T) = 0$ and $\lim_{T\to 0} G_{\varepsilon,D}(x,y,0,T) = \delta(y-x)$. From a diffusion process point of view: In Harrison [19, p. 11-12], for a diffusion process X_t satisfying a SDE $dX_t = \mu dt + \sigma dB_t$, $X_0 = 0$, by the method of images and change of measure,*

$$P(X_t \in dx, M_t \le y) = f_t(x,y)dx,$$

where

$$f_t(x,y) = \frac{1}{\sigma} \exp\left(\frac{\mu x}{\sigma^2} - \frac{\mu^2 t}{2\sigma^2} \right) g_t\left(\frac{x}{\sigma}, \frac{y}{\sigma} \right),$$

$$g_t(x,y) = \left[\phi(x/\sqrt{t}) - \phi((x-2y)/\sqrt{t}) \right] \cdot t^{1/2},$$

and $M_t = \max_{0\le s\le t} X_s$, $\phi(\cdot)$ is the probability density function for a standard normal distribution. By replacing x by $y-x$, μ by 1, and σ by $\sqrt{\varepsilon}$ for our case, the formula for G_ε as in (2.46) can be derived.

More generally, the Dirichlet Green's function for the process $dX_t = \mu dt + \sqrt{\varepsilon} dB_t$ is

$$G_{\varepsilon,D}(x,y,0,t) = \frac{1}{\sqrt{2\pi\varepsilon t}} \left[\exp\left(-\frac{(x-y-\mu t)^2}{2\varepsilon t} \right) - \exp\left(\frac{2\mu x}{\varepsilon} \right) \exp\left(-\frac{(x+y+\mu t)^2}{2\varepsilon t} \right) \right]. \tag{2.47}$$

From (2.41) with $A(\cdot) \equiv 0$ and (2.46),

$$\frac{G_{\varepsilon,D}(x,y,0,T)}{G_\varepsilon(x,y,0,T)} = 1 - \exp[-2xy/\varepsilon T]. \tag{2.48}$$

This can be interpreted in terms of conditional probability for the solution $Y_\varepsilon(s)$, $s \geq 0$ of (2.31) with $b(\cdot,\cdot) \equiv -1$,

$$P\left(\inf_{0\leq s\leq T} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(0) = y, Y_\varepsilon(T) = x\right) = 1 - \exp[-2xy/\varepsilon T]. \tag{2.49}$$

We hope to generalize the result (2.49) to the case $b(y,t) = A(t)y - 1$ in a way that is uniform as $\varepsilon \to 0$.

To further characterize $G_{\varepsilon,D}(x,y,0,T)$ in the case $b(y,t) = A(t)y - 1$, we would like to know more about $Y_\varepsilon(\cdot)$ as defined in (2.31). First, we connect it with a classical control problem

$$q(x,y,t,T) = \min_{y(\cdot)} \left\{ \frac{1}{2}\int_t^T \left[\frac{dy(s)}{ds} - b(y(s),s)\right]^2 ds \middle| y(t) = y, y(T) = x \right\}. \tag{2.50}$$

The Euler-Lagrange equation for the minimizing trajectory $y(\cdot)$ of (2.50) is

$$\frac{d}{ds}\left[\frac{dy(s)}{ds} - b(y(s),s)\right] + \frac{\partial b}{\partial y}(y(s),s)\left[\frac{dy(s)}{ds} - b(y(s),s)\right] = 0, \quad t \leq s \leq T, \tag{2.51}$$

with boundary conditions $y(t) = y, y(T) = x$.

In the case $b(y,t) = A(t)y - 1$, this equation becomes

$$\left[-\frac{d^2}{ds^2} + A'(s) + A(s)^2\right] y(s) = A(s), \quad t \leq s \leq T, \tag{2.52}$$

Let $v(s) = \frac{dy(s)}{ds} - b(y(s),s)$. Taking $t = 0$, then

$$\frac{dv}{ds} + A(s)v(s) = 0, \quad 0 \leq s \leq T,$$

The solution is

$$v(s) = v(T)\exp\left(\int_s^T A(s')ds'\right). \tag{2.53}$$

By solving

$$\frac{dy}{ds} - b(y(s), s) = v(s),$$

we obtain

$$\frac{d}{ds}\left[\exp\left(\int_s^T A(s')ds'\right)y(s)\right] = \exp\left(\int_s^T A(s')ds'\right)[v(s)-1].$$

Since $y(T) = x, y(0) = y$, we have

$$\int_0^T \exp\left(\int_s^T A(s')ds'\right)v(s)ds = x + m_2(T) - m_1(T)y.$$

Together with (2.53), we get

$$v(T) = \frac{x + m_2(T) - m_1(T)y}{\sigma^2(T)}.$$

Thus,

$$v(s) = \frac{dy(s)}{ds} - b(y(s), s) = \frac{x + m_2(T) - m_1(T)y}{\sigma^2(T)}\exp\left(\int_s^T A(s')ds'\right). \tag{2.54}$$

Plugging $v(s)$ into (2.50), we obtain the formula for $q(x, y, 0, T)$,

$$q(x, y, 0, T) = \frac{(x + m_2(T) - m_1(T)y)^2}{2\sigma^2(T)}. \tag{2.55}$$

Therefore, the connection between the Green's function for (2.31) and the control problem (2.50) in the case $b(y, t) = A(t)y - 1$ is given by

$$G_\varepsilon(x, y, 0, T) = \frac{1}{\sqrt{2\pi\varepsilon\sigma^2(T)}}\exp[-q(x, y, 0, T)/\varepsilon]. \tag{2.56}$$

The minimizing trajectory $y(\cdot)$ for (2.50) has probabilistic significance as well as the function $q(x, y, t, T)$. The probability density function for $Y_\varepsilon(s) = z$ conditioned on $Y_\varepsilon(T) = x$ is

$$\frac{G_\varepsilon(z,y,t,s)\cdot G_\varepsilon(x,z,s,T)}{G_\varepsilon(x,y,t,T)} = \frac{1}{\sqrt{2\pi\varepsilon\frac{\sigma^2(t,s)\sigma^2(s,T)}{\sigma^2(t,T)}}}\exp\left[-\frac{(z-y(s))^2}{2\varepsilon\frac{\sigma^2(t,s)\sigma^2(s,T)}{\sigma^2(t,T)}}\right],$$

where

$$y(s) = \frac{1}{\sigma^2(t,T)}\left[xm_1(s,T)\sigma^2(t,s) + ym_1(t,s)\sigma^2(s,T) + \right.$$
$$\left. m_1(s,T)m_2(s,T)\sigma^2(t,s) - m_2(t,s)\sigma^2(s,T)\right] \quad (2.57)$$

Therefore, the solution $Y_\varepsilon(s)$, $0 \le s \le T$ of (2.31) conditioned on $Y_\varepsilon(0) = y, Y_\varepsilon(T) = x$ is a Gaussian process with mean and variance at s given by

$$E[Y_\varepsilon(s)|Y_\varepsilon(0) = y, Y_\varepsilon(T) = x] = y(s), \quad 0 \le s \le T, \quad (2.58)$$

$$Var[Y_\varepsilon(s)|Y_\varepsilon(0) = y, Y_\varepsilon(T) = x] = \varepsilon\sigma^2(0,s)\sigma^2(s,T)/\sigma^2(0,T), \quad (2.59)$$

where $y(s)$ is defined in (2.57) with $t = 0$. Also, by examining

$$G_\varepsilon(z_1,y,t,s_1)G_\varepsilon(z_2,z_1,s_1,s_2)G_\varepsilon(x,z_2,s_2,T)/G_\varepsilon(x,y,t,T),$$

we can obtain the covariance of $Y_\varepsilon(\cdot)$ conditioned on $Y_\varepsilon(T) = x$ and $Y_\varepsilon(t) = y$:

$$\text{Covar}[Y_\varepsilon(s_1), Y_\varepsilon(s_2)|Y_\varepsilon(t) = y, Y_\varepsilon(T) = x] = \varepsilon\Gamma(s_1,s_2), \quad 0 \le s_1 \le s_2 \le T, \quad (2.60)$$

where the symmetric function $\Gamma : [t,T]\times[t,T] \to \mathrm{R}$ is given by the formula

$$\Gamma(s_1,s_2) = \frac{m_1(s_1,s_2)\sigma^2(t,s_1)\sigma^2(s_2,T)}{\sigma^2(t,T)}. \quad (2.61)$$

We now show the relation between the conditioned process $Y_\varepsilon(s)$ and the control problem (2.50). By solving (2.54), we have

$$y(s) = \frac{x - (x + m_2(T) - m_1(T)y) \cdot \frac{\sigma^2(s,T)}{\sigma^2(T)} + m_2(s,T)}{m_1(s,T)}$$

Since

$$\begin{aligned}
\frac{1 - \sigma^2(s,T)/\sigma^2(T)}{m_1(s,T)} &= \frac{\sigma^2(T) - \sigma^2(s,T)}{m_1(s,T)\sigma^2(T)} = \frac{m_1(s,T)\sigma^2(0,s)}{\sigma^2(T)}, \\
\frac{m_1(T)\sigma^2(s,T)}{m_1(s,T)\sigma^2(T)} &= \frac{m_1(0,s)\sigma^2(s,T)}{\sigma^2(T)}, \\
\frac{-m_2(T) \cdot \sigma^2(s,T)/\sigma^2(T) + m_2(s,T)}{m_1(s,T)} &= \frac{m_2(s,T)(\sigma^2(0,s)m_1^2(s,T) + \sigma^2(s,T)) - m_2(T)\sigma^2(s,T)}{m_1(s,T)\sigma^2(T)} \\
&= \frac{m_1(s,T)m_2(s,T)\sigma^2(0,s) - m_2(0,s)\sigma^2(s,T)}{\sigma^2(T)},
\end{aligned}$$

(in deriving this identities, we need to use identities (2.42), (2.43)) we have

$$y(s) = x\frac{m_1(s,T)\sigma^2(0,s)}{\sigma^2(T)} + y\frac{m_1(0,s)\sigma^2(s,T)}{\sigma^2(T)} + \frac{m_1(s,T)m_2(s,T)\sigma^2(0,s) - m_2(0,s)\sigma^2(s,T)}{\sigma^2(T)}. \tag{2.62}$$

Therefore, the optimal trajectory of (2.50) is the same as the expectation of $Y_\varepsilon(s)$ given $Y_\varepsilon(0) = y, Y_\varepsilon(T) = x$ in (2.57).

2.3.3 A non-Markovian representation

In this subsection, we derive a non-Markovian representation for the process $Y_\varepsilon(\cdot)$ with given initial and terminal conditions. First we note that the function Γ as defined in (2.61) is the Dirichlet Green's function for the operator on the LHS of (2.52). Thus taking $t = 0$, one has

$$\left[-\frac{d^2}{ds_1^2} + A'(s_1) + A(s_1)^2\right]\Gamma(s_1, s_2) = \delta(s_1 - s_2), \quad 0 < s_1, s_2 < T, \tag{2.63}$$

and $\Gamma(0, s_2) = \Gamma(T, s_2) = 0$ for all $0 < s_2 < T$. By exploiting this fact, we can construct the non-Markovian representation for $Y_\varepsilon(\cdot)$ that we need.

Remark II.12. *To show (2.63), when $s_1 < s_2$, (the proof for $s_1 > s_2$ is similar so omitted.)*

$$\begin{aligned}
-\frac{d^2}{ds_1^2}\Gamma(s_1, s_2) = &-\frac{\sigma^2(s_2, T)}{\sigma^2(T)} \cdot \bigg(\frac{d^2}{ds_1^2} m_1(s_1, s_2)\sigma^2(0, s_1) \\
&+ 2\frac{d}{ds_1} m_1(s_1, s_2)\frac{d}{ds_1}\sigma^2(0, s_1) + \frac{d^2}{ds_1^2}\sigma^2(0, s_1) m_1(s_1, s_2)\bigg) \\
= &-\frac{\sigma^2(s_2, T)}{\sigma^2(T)} \cdot \big((A^2(s_1) - A'(s_1)) m_1(s_1, s_2)\sigma^2(0, s_1) \\
&- 2A(s_1) m_1(s_1, s_2) \cdot (1 + 2A(s_1)\sigma^2(0, s_1)) \\
&+ 2A'(s_1)\sigma^2(0, s_1) m_1(s_1, s_2) + 2A(s_1)(1 + 2A(s_1)\sigma^2(0, s_1)) m_1(s_1, s_2)\big) \\
= &-\frac{\sigma^2(s_2, T)}{\sigma^2(T)} \cdot (A^2(s_1) + A'(s_1)) m_1(s_1, s_2)\sigma^2(0, s_1)
\end{aligned}$$

thus

$$\left[-\frac{d^2}{ds_1^2} + A'(s_1) + A(s_1)^2\right]\Gamma(s_1, s_2) = 0, \quad \text{for } s_1 < s_2.$$

We notice that $\partial\Gamma(s_1, s_2)/\partial s_1$ is not continuous around $s_1 = s_2$. Actually, when $s_1 \uparrow s_2$,

$$\lim_{s_1 \to s_2^-} \frac{\partial\Gamma(s_1, s_2)}{\partial s_1} = \frac{\sigma^2(s, T)}{\sigma^2(T)}\left[A(s)\sigma^2(0, s) + 1\right],$$

where $s = s_2$. when $s_1 \downarrow s_2$,

$$\lim_{s_1 \to s_2^+} \frac{\partial\Gamma(s_1, s_2)}{\partial s_1} = \frac{\sigma^2(0, s)}{\sigma^2(T)}\left[A(s)\sigma^2(s, T) - m_1^2(s, T).\right]$$

Thus

$$\left(\lim_{s_1 \to s_2^-} - \lim_{s_1 \to s_2^+}\right)\frac{\partial\Gamma(s_1, s_2)}{\partial s_1} = \frac{\sigma^2(s, T) + \sigma^2(0, s) m_1^2(s, T)}{\sigma^2(0, T)} = 1,$$

where the second equality holds due to (2.42). Therefore (2.63) is true.

Next, we try to derive a representation of the conditioned process $Y_\varepsilon(\cdot)$ in terms of the white noise process, by obtaining a factorization of Γ corresponding to the factorization

$$-\frac{d^2}{ds^2} + A'(s) + A(s)^2 = \left[-\frac{d}{ds} - A(s)\right]\left[\frac{d}{ds} - A(s)\right].$$

We note that the boundary value problem

$$\left[\frac{d}{ds} - A(s)\right] u(s) = v(s), \quad 0 < s < T, \quad u(0) = u(T) = 0, \tag{2.64}$$

has a solution if and only if the function $v : [0, T] \to \mathrm{R}$ satisfies the orthogonality condition

$$\int_0^T \frac{v(s)}{m_1(0,s)} ds = 0. \tag{2.65}$$

Therefore, in order to solve the boundary value problem

$$\left[-\frac{d^2}{ds^2} + A'(s) + A(s)^2\right] u(s) = f(s), \quad 0 < s < T, \quad u(0) = u(T) = 0, \tag{2.66}$$

we only need to find the solution u to

$$\left[-\frac{d}{ds} - A(s)\right] v(s) = f(s), \quad 0 < s < T, \tag{2.67}$$

which satisfies the orthogonality condition (2.65). The solution to (2.65), (2.67) is given by an expression

$$v(s) = K^* f(s) := \int_0^T k(s', s) f(s') ds', \quad 0 \le s \le T, \tag{2.68}$$

where the kernel $k : [0,T] \times [0,T] \to \mathrm{R}$ is defined by

$$k(s', s) = \frac{m_1(s, s')\sigma^2(s', T)}{\sigma^2(T)} \text{ if } s' > s,$$

$$k(s', s) = \frac{\sigma^2(s', T)}{m_1(s', s)\sigma^2(T)} - \frac{1}{m_1(s', s)} \text{ if } s' < s. \quad (2.69)$$

If v satisfies the condition (2.65), then

$$u(s) = Kv(s) = \int_0^T k(s, s')v(s')ds', \quad 0 \leq s \leq T, \quad (2.70)$$

is the solution to (2.64).

Remark II.13. *In this remark, we show the results stated above. By solving (2.67), we obtain*

$$v(s) = \frac{v(0)}{m_1(0, s)} - \int_0^s \frac{f(s')}{m_1(s', s)} ds'.$$

Due to (2.65), we have

$$\begin{aligned} v(0) &= \int_0^T \frac{ds}{m_1(0, s)} \int_0^s \frac{f(s')ds'}{m_1(s', s)} \Big/ \int_0^T \frac{1}{m_1^2(0, s)} ds \\ &= \int_0^T ds' f(s') l(s', s), \end{aligned}$$

where

$$l(s') = \int_{s'}^T \frac{ds''}{m_1(0, s'') \cdot m_1(s', s'')} \Big/ \int_0^T \frac{1}{m_1^2(0, s'')} ds''.$$

It is easy to show (2.68), (2.69) from here.

As for the expression of $u(s)$, by solving (2.64), we have

$$u(s) = m_1(s) \int_0^s \frac{v(s')}{m_1(s')} ds' = -m_1(s) \int_s^T \frac{v(s')}{m_1(s')} ds'.$$

Thus, $u(s)$ can as well be expressed as

$$
\begin{aligned}
u(s) &= \left[\left(\frac{\sigma^2(s,T)m_1(s)}{\sigma^2(T)}\right)\int_0^s \frac{v(s')}{m_1(s')}ds' - \left(m_1(s) - \frac{\sigma^2(s,T)m_1(s)}{\sigma^2(T)}\right)\int_s^T \frac{v(s')}{m_1(s')}ds'\right] \\
&= \int_0^s \frac{m_1(s',s)\sigma^2(s,T)}{\sigma^2(T)}v(s')ds' + \int_s^T \left[\frac{\sigma^2(s,T)}{m_1(s,s')\sigma^2(T)} - \frac{1}{m_1(s,s')}\right]v(s')ds' \\
&= \int_0^T k(s,s')v(s')ds'.
\end{aligned}
$$

It follows that the kernel Γ, as an operator, has the factorization $\Gamma = KK^*$, where

$$
Kf(s) = \int_0^T k(s,s')f(s')ds', \quad K^*f(s) = \int_0^T k(s',s)f(s')ds',
$$

thereby

$$
u(s) = \Gamma f(s) = \int_0^T \Gamma(s,s')f(s')ds' = \int_0^T k(s,s')\left(\int_0^T k(s'',s')f(s'')ds''\right)ds' = KK^*f(s),
$$

and

$$
\Gamma(s_1,s_2) = \int_0^T k(s_1,s)k(s_2,s)ds.
$$

Therefore, the conditioned process $Y_\varepsilon(\cdot)$ has the representation

$$
Y_\varepsilon(s) = y(s) + \sqrt{\varepsilon}\int_0^T k(s,s')dB(s'), \quad 0 \le s \le T. \tag{2.71}
$$

Remark II.14. *It is easy to check that the covariance of this representation (2.71) is the same as $\Gamma(s_1,s_2)$:*

$$
E\left[\sqrt{\varepsilon}\int_0^T k(s_1,s')dB(s')\sqrt{\varepsilon}\int_0^T k(s_2,s')dB(s')\right] = \varepsilon\int_0^T k(s_1,s')k(s_2,s')ds' = \Gamma(s_1,s_2).
$$

When $A(\cdot) \equiv 0$, equation (2.71) yields the familiar representation

$$Y_\varepsilon(s) = \frac{s}{T}x + \left(1 - \frac{s}{T}\right) y + \sqrt{\varepsilon}\left[B(s) - \frac{s}{T}B(T)\right], \quad 0 \le s \le T,$$

for the Brownian bridge process.

2.3.4 A Markovian representation

We notice that the representation (2.71) of the conditioned process $Y_\varepsilon(s)$ is not Markovian. In the following, we obtain an alternative representation by considering a stochastic control problem. Let $Y_\varepsilon(\cdot)$ be the solution to the stochastic differential equation

$$dY_\varepsilon(s) = \lambda_\varepsilon(\cdot, s)ds + \sqrt{\varepsilon}dB(s), \tag{2.72}$$

where $\lambda_\varepsilon(\cdot, s)$ is a non-anticipating function. We consider the problem of minimizing the cost function given by the formula

$$q_\varepsilon(x, y, t, T) = \min_{\lambda_\varepsilon} E\left[\frac{1}{2}\int_t^T [\lambda_\varepsilon(\cdot, s) - b(Y_\varepsilon(s), s)]^2 ds \middle| Y_\varepsilon(t) = y, Y_\varepsilon(T) = x\right]. \tag{2.73}$$

The minimum is to be taken over all non-anticipating $\lambda_\varepsilon(\cdot, s)$, $t \le s < T$, which have the property that the solution of (2.72) with initial condition $Y_\varepsilon(t) = y$ satisfy the terminal condition $Y_\varepsilon(T) = x$ with probability 1. The optimal controller λ^* for the problem is given formally by the expression

$$\lambda_\varepsilon(\cdot, s) = \lambda^*_\varepsilon(x, Y_\varepsilon(s), s) = b(Y_\varepsilon(s), s) - \frac{\partial q_\varepsilon}{\partial y}(x, Y_\varepsilon(s), s),$$

where q_ε satisfies the HJB equation

$$
\begin{aligned}
0 &= \min_{\lambda}\left[\frac{1}{2}(\lambda - b(y,t))^2 + \frac{\partial q_\varepsilon}{\partial y}\lambda + \frac{\varepsilon}{2}\frac{\partial^2 q_\varepsilon}{\partial y^2} + \frac{\partial q_\varepsilon}{\partial t}\right] \\
&= \frac{\varepsilon}{2}\frac{\partial^2 q_\varepsilon}{\partial y^2} - \frac{1}{2}\left(\frac{\partial q_\varepsilon}{\partial y}\right)^2 + b(y,t)\frac{\partial q_\varepsilon}{\partial y} + \frac{\partial q_\varepsilon}{\partial t}.
\end{aligned}
$$

In the classical case, $\varepsilon = 0$, the solution to (2.72), (2.73) is the same as the variational problem (2.50). When $b(y,t)$ is linear in y, the problem is of linear-quadratic type and the difference between the cost functions for the classical and stochastic control problems is independent of y, therefore,

$$
\lambda_\varepsilon^*(x,y,t) = b(y,t) - \frac{\partial q(x,y,t,T)}{\partial y} = A(t)y - 1 - \frac{\partial}{\partial y}\frac{[x + m_2(t,T) - m_1(t,T)y]^2}{2\sigma^2(t,T)}. \tag{2.74}
$$

It is easy to see that if we solve the SDE (2.72) with controller given by (2.74) and conditioned on $Y_\varepsilon(t) = y$, then $Y_\varepsilon(T) = x$ with probability 1. In fact, this Markovian process $Y_\varepsilon(s)$, $t \le s \le T$ has the same distribution as the process $Y_\varepsilon(s)$, $t \le s \le T$, satisfying the SDE (2.31) conditioned on $Y_\varepsilon(t) = y$, $Y_\varepsilon(T) = x$.

Remark II.15. *In this remark, we justify the statement in the end of the previous paragraph. Substituting $\lambda_\varepsilon(\cdot, s)$ by the optimal choice in (2.74), we have the SDE for $Y_\varepsilon(s)$ as*

$$
dY_\varepsilon(s) + \left[\frac{m_1^2(s,T)}{\sigma^2(s,T)} - A(s)\right]Y_\varepsilon(s)ds = \left(-1 + \frac{m_1(s,T)(x + m_2(s,T))}{\sigma^2(s,T)}\right)ds + \sqrt{\varepsilon}dB(s). \tag{2.75}
$$

Multiplying both sides by the integral factor $m_1(s,T)/\sigma^2(s,T)$, it follows that

$$
d\left(\frac{m_1(s,T)}{\sigma^2(s,T)}Y_\varepsilon\right) = \frac{m_1(s,T)}{\sigma^2(s,T)}\left(-1 + \frac{m_1(s,T)(x + m_2(s,T))}{\sigma^2(s,T)}\right)ds + \sqrt{\varepsilon}\frac{m_1(s,T)}{\sigma^2(s,T)}dB.
$$

Since

$$\int_0^s \frac{m_1(s',T)}{\sigma^2(s',T)}\left(-1+\frac{m_1(s',T)(x+m_2(s',T))}{\sigma^2(s',T)}\right)ds' =$$
$$x\left[\frac{1}{\sigma^2(s,T)}-\frac{1}{\sigma^2(T)}\right]-\frac{m_2(0,T)}{\sigma^2(0,T)}+\frac{m_2(s,T)}{\sigma^2(s,T)},$$

the mean of $Y_\varepsilon(s)$ *is*[2]

$$E[Y_\varepsilon(s)] = x\frac{\sigma^2(0,s)m_1(s,T)}{\sigma^2(T)}+y\frac{m_1(0,s)\sigma^2(s,T)}{\sigma^2(T)}+$$
$$\frac{1}{\sigma^2(T)}\left[m_1(s,T)m_2(s,T)\sigma^2(0,s)-m_2(0,s)\sigma^2(s,T)\right],$$

the same as (2.57)*. We can see that the variance of* $Y_\varepsilon(s)$ *is*

$$\varepsilon\int_0^s \frac{m_1^2(s',T)}{[\sigma^2(s',T)]^2}ds' \times \left[\frac{\sigma^2(s,T)}{m_1(s,T)}\right]^2 = \varepsilon\frac{1}{\sigma^2(s',T)}\Big|_0^s \times \left[\frac{\sigma^2(s,T)}{m_1(s,T)}\right]^2 = \varepsilon\frac{\sigma^2(0,s)\sigma^2(s,T)}{\sigma^2(T)},$$

the same as (2.59)*. As for the covariance, if* $0 \le s_1 \le s_2 \le T$

$$\begin{aligned} E[(Y_\varepsilon(s_1)-y(s_1))(Y_\varepsilon(s_2)-y(s_2)] &= \varepsilon\frac{\sigma^2(s_1,T)\sigma^2(s_2,T)}{m_1(s_1,T)m_1(s_2,T}\int_0^{s_1}\left[\frac{m_1(s',T)}{\sigma^2(s',T)}\right]^2 ds' \\ = \varepsilon\frac{\sigma^2(s_1,T)\sigma^2(s_2,T)}{m_1(s_1,T)m_1(s_2,T}\left[-\frac{1}{\sigma^2(s',T)}\right]\Bigg|_0^{s_1} &= \varepsilon\frac{\sigma^2(s_1,T)\sigma^2(s_2,T)}{m_1(s_1,T)m_1(s_2,T}\left[\frac{1}{\sigma^2(T)}-\frac{1}{\sigma^2(s_1,T)}\right] \\ &= \varepsilon\frac{\sigma^2(0,s_1)m_1(s_1,s_2)\sigma^2(s_2,T)}{\sigma^2(T)}, \quad (2.76) \end{aligned}$$

which is the same as $\varepsilon\Gamma(s_1,s_2)$, $\Gamma(s_1,s_2)$ *being defined in* (2.61)*. In the last step of* (2.76)*, we use* (2.42)*. Therefore, the Markovian process satisfying* (2.75) *is the same as the solution* $Y_\varepsilon(s)$, $0 \le s \le T$ *of* (2.31) *conditioned on* $Y_\varepsilon(0)=y, Y_\varepsilon(T)=x$ *with* $b(y,s)=A(s)y-1$.

Remark II.16. *We note that* q_ε *is logarithmically divergent at* $s=T$ *with the optimal*

[2]It should be noticed that $m_2(s,T)-m_2(0,T)=-m_1(s,T)m_2(0,s)$.

controller: With λ_ε^* chose in (2.74), the function $q_\varepsilon(x,y,t,T)$ is approximately

$$E\left[\frac{1}{2}\int_t^T\left(\frac{x-Y_\varepsilon(s)}{T-s}\right)^2 ds\right]. \tag{2.77}$$

Some other terms are neglected for the purpose of demonstration. When s is close to T, $Y_\varepsilon(s)$ is governed by an approximated SDE

$$dY_\varepsilon(s)=\frac{x-Y_\varepsilon(s)}{T-s}ds+\sqrt{\varepsilon}dB(s),$$

thus

$$d\left[\frac{Y_\varepsilon(s)-x}{T-s}\right]=\frac{\sqrt{\varepsilon}}{T-s}dB(s),$$

which further implies

$$\frac{Y_\varepsilon(s)-x}{T-s}=\frac{y-x}{T}+\sqrt{\varepsilon}\int_0^s\frac{1}{T-s}dB(s).$$

Therefore, it follows (2.77) that

$$q_\varepsilon(x,y,t,T)\approx\frac{\varepsilon}{2}\int_t^T E\left(\int_0^s\frac{1}{T-s'}dB(s')\right)^2 ds=\frac{\varepsilon}{2}\int_t^T\left(\frac{1}{T-s}-\frac{1}{T}\right)ds.$$

We notice that $\int_t^T 1/(T-s)ds$ diverges logarithmically. Thus q_ε is not well-defined for all when $t=T$. However, $\partial q_\varepsilon/\partial y$ always exists and is continuous.

Solving (2.72) with drift (2.74) and $Y_\varepsilon(0)=y$, then if $t=0$ (2.71) holds with another kernel k given by

$$k(s,s')=\frac{m_1(s',s)\sigma^2(s,T)}{\sigma^2(s',T)}\quad\text{if } s'<s,\quad k(s,s')=0\quad\text{if } s'>s. \tag{2.78}$$

This kernel corresponds to the Cholesky factorization $\Gamma=KK^*$ for the kernel Γ.

In the case $A(\cdot) \equiv 0$ equation (2.71) yields the Markovian representation

$$Y_\varepsilon(s) = \frac{s}{T}x + \left(1 - \frac{s}{T}\right) y + \sqrt{\varepsilon}(T-s) \int_0^s \frac{dB(s')}{T-s'}, \quad 0 \le s \le T, \tag{2.79}$$

for the Brownian bridge process.

With the help of the Markovian representation (2.72), we can express the ratio in (2.48) of Green's functions for the linear case $b(y,t) = A(t)y - 1$ in terms of the solution to a partial differential equation. We assume $x > 0$ and define

$$u(y,t) = P\left(\inf_{t \le s \le T} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(t) = y \right), \quad y > 0, t < T, \tag{2.80}$$

where $Y_\varepsilon(\cdot)$ is the solution to the SDE (2.72) with drift (2.74). Then $u(y,t)$ is the solution to the PDE

$$\frac{\partial u(y,t)}{\partial t} + \lambda_\varepsilon^*(x,y,t)\frac{\partial u(y,t)}{\partial y} + \frac{\varepsilon}{2}\frac{\partial^2 u(y,t)}{\partial y^2} = 0, \quad y > 0, t < T, \tag{2.81}$$

with boundary and terminal conditions

$$u(0,t) = 0 \text{ for } t < T, \quad \lim_{t \to T} u(y,t) = 1 \text{ for } y > 0. \tag{2.82}$$

In the case $A(\cdot) \equiv 0$, the PDE (2.81) becomes

$$\frac{\partial u(y,t)}{\partial t} + \left(\frac{x-y}{T-t}\right)\frac{\partial u(y,t)}{\partial y} + \frac{\varepsilon}{2}\frac{\partial^2 u(y,t)}{\partial y^2} = 0, \quad y > 0, t < T. \tag{2.83}$$

It can easily be shown that

$$u(y,t) = 1 - \exp\left[-\frac{2xy}{\varepsilon(T-t)}\right], \quad t < T, y > 0$$

is the solution to (2.82), (2.83). We note that when $t = 0$, the function above is the same as (2.48).

Remark II.17. *A drawback of the Markovian representation is that the contribution from the BM part is infinite towards the end $t = T$. (See (2.79).)*

2.3.5 Estimation on the Dirichlet Green's function

The Dirichlet Green's function has a crucial role in deriving the classic limit theorem. In this section, we study the relationship between the full space Green's function (without a boundary) and the Dirichlet Green's function in a more general setting. We know that the ratio of the Dirichlet Green's function to the full space Green's function, $G_{\varepsilon,D}(x, y, 0, T)/G_\varepsilon(x, y, 0, T)$, is the same as the probability

$$P\left(\inf_{0\le s\le T} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(0) = y, Y_\varepsilon(T) = x\right),$$

which has an explicit expression when $A(\cdot) \equiv 0$. (See (2.48), (2.49).) We will show similar results for the non-zero case of $A(\cdot)$ when $\varepsilon \to 0$.

Proposition II.18. *Assume $b(y, t) = A(t)y - 1$ where (2.27) holds and the function $A(\cdot)$ is non-negative. Then for $\lambda, y, T > 0$ the ratio of the Dirichlet to full space Green's function satisfies the limit*

$$\lim_{\varepsilon\to 0} \frac{G_{\varepsilon,D}(\lambda\varepsilon, y, 0, T)}{G_\varepsilon(\lambda\varepsilon, y, 0, T)} = 1 - \exp\left[-2\lambda\left\{1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right\}\right]. \tag{2.84}$$

Remark II.19. *We note that for a process*

$$dZ_\varepsilon(t) = \mu dt + \sqrt{\varepsilon}dB(t), \quad Z_\varepsilon(0) = \lambda\varepsilon, \tag{2.85}$$

with $\mu > 0$, we have

$$P\left(\inf_{t>0} Z_\varepsilon(t) < 0\right) = e^{-2\lambda\mu}\ {}^{3}. \tag{2.86}$$

[3]Show that $P(\inf_{t>0} Z_\varepsilon(t) < 0) = e^{-2\lambda\mu}$: Suppose that $f(x) = P(\inf_{t>0} Z_\varepsilon(t) > 0|Z_\varepsilon(0) = \lambda\varepsilon)$.

The right hand side of (2.84) *behaves like the probability of* $\inf_{t>0} Z_\varepsilon(t)$ *being positive with* $\mu = 1 - m_2(T)/\sigma^2(T) + m_1(T)y/\sigma^2(T)$. *The ratio on the left hand side of* (2.84), $G_{\varepsilon,D}(\lambda\varepsilon, y, 0, T)/G_\varepsilon(\lambda\varepsilon, y, 0, T)$, *is the probability for* $Y_\varepsilon(s)$ *to be positive over* $[0, T]$, *where* $Y_\varepsilon(s)$ *is defined as in* (2.71) *with* x *in* (2.57) *being* $\lambda\varepsilon$.

Since $Y_\varepsilon(s)$ *is most likely to exit through the boundary 0 at a time* $T - O(\varepsilon)$[4], *as* $\varepsilon \to 0$, *only the drift of* $Y_\varepsilon(s)$ *close to the time point* T *matters to the probability for* $Y_\varepsilon(s) > 0$. *This drift is* $1 - m_2(T)/\sigma^2(T) + m_1(T)y/\sigma^2(T)$, *which intuitively explains why the* μ *in* (2.86) *is substituted by this specific value. We note that under the assumption that* $A(\cdot) \geq 0$, *this drift is positive. Actually, it can be easily shown by noting*

$$\frac{m_2(T)}{\sigma^2(T)} = \frac{\int_0^T \exp(\int_s^T A(s')ds')ds}{\int_0^T \exp(\int_s^T 2A(s')ds')ds} \leq 1.$$

Proof of Proposition II.18. According to (2.71), let

$$Y_\varepsilon(s) = y(s) + \sqrt{\varepsilon}\left[\frac{m_1(s)\sigma^2(s,T)}{\sigma^2(T)}\int_0^T \frac{dB(s')}{m_1(s')} - m_1(s)\int_s^T \frac{dB(s')}{m_1(s')}\right], \tag{2.87}$$

with $Y_\varepsilon(T) = \lambda\varepsilon$. In order to find $\lim_{\varepsilon\to 0} P(\inf_{0\leq s\leq T} Y_\varepsilon(s) > 0)$, we derive both upper and lower bounds for this probability.

Before the analysis on the two bounds, we first have a close look at the derivative of the mean of the random process $Y_\varepsilon(s)$, $y(s)$. Following (2.57), we have

$$\begin{aligned} y'(s) = \frac{1}{\sigma^2(T)}\Big\{ &\lambda\varepsilon \cdot m_1(s,T)(1 + A(s)\sigma^2(0,s)) \\ &+ y \cdot m_1(0,s)[A(s)\sigma^2(s,T) - m_1^2(s,T)] \\ &+ m_1(s,T)m_2(s,T)(1 + A(s)\sigma^2(0,s)) - m_1^2(s,T)\sigma^2(0,s)) \\ &- (1 + A(s)m_2(0,s))\sigma^2(s,T) + m_2(0,s)m_1^2(s,T)\Big\}. \end{aligned} \tag{2.88}$$

Since $dZ_\varepsilon(t) = \mu dt + \sqrt{\varepsilon}dB(t)$, $f(x)$ satisfies a differential equation $\mu f'(x) + \varepsilon f''(x)/2 = 0$, $f(0) = 1$. The solution of this differential equation with condition $\lim_{x\to\infty} f(x) = 0$ is $e^{-2\lambda\mu}$.

[4] Suppose the exiting time is τ, then $\mu\tau + \sqrt{\varepsilon}dB(\tau) = -\lambda\varepsilon$. Since $B(\tau) \sim \sqrt{\tau}$, as $\varepsilon \to 0$, $\tau = O(\varepsilon)$.

At $s = T$,

$$y'(T) = O(\varepsilon) - 1 + \frac{m_2(T)}{\sigma^2(T)} - \frac{m_1(T)y}{\sigma^2(T)}. \tag{2.89}$$

For the upper bound, as long as $a\varepsilon < T$, where a is a variable relying on ε to be specified later, we have

$$P\left(\inf_{0\le s\le T} Y_\varepsilon(s) > 0\right) \le P\left(\inf_{0\le t\le a\varepsilon} Y_\varepsilon(T-t) > 0\right) = P\left(\inf_{0\le t\le a\varepsilon}[Z_\varepsilon(t) + \tilde{Z}_\varepsilon(t)] > 0\right), \tag{2.90}$$

where $Z_\varepsilon(t)$ is a stochastic process adopting a fixed drift which is the same as $-y'(T)$ without the ε-related part:

$$dZ_\varepsilon(t) = \mu dt + \sqrt{\varepsilon} dB(t), \quad Z_\varepsilon(0) = \lambda\varepsilon, \tag{2.91}$$

where $\mu = 1 - m_2(T)/\sigma^2(T) + m_1(T)y/\sigma^2(T)$, and from (2.87),

$$\begin{aligned}&\tilde{Z}_\varepsilon(t) = Y_\varepsilon(T-t) - Z_\varepsilon(t) = y(T-t) - y(T) + y'(T)t\\&+\sqrt{\varepsilon}\left[\frac{m_1(T-t)\sigma^2(T-t,T)}{\sigma^2(T)}\int_0^T \frac{dB(s')}{m_1(s')} + \int_{T-t}^T \left(1 - \frac{m_1(T-t)}{m_1(s')}\right) dB(s')\right].\end{aligned} \tag{2.92}$$

Here $Z_\varepsilon(t)$ is the "linearization" of $Y_\varepsilon(T-t)$ and $\tilde{Z}_\varepsilon(t)$ is the "error" of the approximation.

We use the inequality

$$P\left(\inf_{0\le t\le a\varepsilon}[Z_\varepsilon(t) + \tilde{Z}_\varepsilon(t)] > 0\right) \le P\left(\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t) > -b\lambda\varepsilon\right) + P\left(\sup_{0\le t\le a\varepsilon} \tilde{Z}_\varepsilon(t) > b\lambda\varepsilon\right), \tag{2.93}$$

which holds true for any $b > 0$[5].

Since $Z_\varepsilon(t)$ has a constant drift, we are able to estimate the first term on the right

[5] Similarly to a, b will be specifically chosen later. We notice that $\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t)$ is compared with $b\lambda\varepsilon$ since the change of $Z_\varepsilon(t)$ over the interval $[0, a\varepsilon]$ is of the scale $O(\varepsilon)$.

hand side of (2.93) by using *the method of images*[6]:

$$
\begin{aligned}
P\left(\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t) < -b\lambda\varepsilon\right) &= 1-\int_{-b\lambda\varepsilon}^{\infty} P\left(Z_\varepsilon(t)\in dy, \inf_{0\le t\le a\varepsilon} Z_\varepsilon(t) > -b\lambda\varepsilon\right)\\
&= e^{-2\mu(1+b)\lambda}\frac{1}{\sqrt{2\pi}}\int_{[(1+b)\lambda-\mu a]/\sqrt{a}}^{\infty} e^{-z^2/2}dz + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-[(1+b)\lambda+\mu a]/\sqrt{a}} e^{-z^2/2}dz. \quad (2.94)
\end{aligned}
$$

From this equation, we expect the first term gives the main contribution to the bound. We will later choose a and b in a way such that as $\varepsilon \to 0$, the right hand side of (2.94) converges to $\exp\{-2\mu\lambda\}$.

Next we estimate the second term on the right hand side of (2.93). Intuitively, we expect $\tilde{Z}_\varepsilon(a\varepsilon)$ to be smaller than the scale $O(\varepsilon)$. Now we analyze the components of $\tilde{Z}_\varepsilon(t)$ in (2.92) one by one.

1. Based on Taylor expansion,

$$
\sup_{0\le t\le a\varepsilon} |y(T-t)-y(T)+y'(T)t| \le \frac{1}{2}\sup_{0\le t\le a\varepsilon}|y''(t)|a^2\varepsilon^2, \quad 0<a\varepsilon\le T. \quad (2.95)
$$

Due to the expression of $y(s)$ in (2.57), there exists C constant only depending on A, T, such that $y''(t)\le C[\lambda\varepsilon+y+1]$. Therefore

$$
\sup_{0\le t\le a\varepsilon} |y(T-t)-y(T)+y'(T)t| \le C[\lambda\varepsilon+y+1]a^2\varepsilon^2, \quad 0<a\varepsilon\le T. \quad (2.96)
$$

As long as $b\lambda\varepsilon$ converges to 0 slower than $O(\varepsilon^2)$, the contribution of this term to the probability $P\left(\sup_{0\le t\le a\varepsilon}\tilde{Z}_\varepsilon(t) > b\lambda\varepsilon\right)$ is negligible, i.e., when ε is small enough,

$$
P\left(|y(T-t)-y(T)+y'(T)t| > b\lambda\varepsilon/4\right) = 0. \quad (2.97)
$$

[6] See Harrison [19, page 14 (11)]. For a process $dX_t=\mu dt+\sigma dB_t$, $X_0=0$, $P(T(y)>t)=\Phi\left(\frac{y-\mu t}{\sigma t^{1/2}}\right)-e^{2\mu y/\sigma^2}\Phi\left(\frac{-y-\mu t}{\sigma t^{1/2}}\right)$, where $T(y)$ is the hitting time of $y>0$. Here we only need to replace μ by $-\mu$, y by $(b+1)\lambda\varepsilon$, and t by $a\varepsilon$ to obtain the equation.

2. The second term is bounded by

$$\sup_{0\le t\le a\varepsilon}\left|\sqrt{\varepsilon}\frac{m_1(T-t)\sigma^2(T-t,T)}{\sigma^2(T)}\int_0^T\frac{dB(s')}{m_1(s')}\right|.$$

We notice that $m_1(T-t)$ and $\sigma^2(T)$ can both be bounded by constants, and $\sigma^2(T-t,T)=\int_{T-t}^T\exp\left(\int_s^T 2A(s')ds'\right)ds\le\exp(2AT)a\varepsilon$. Thus

$$\sup_{0\le t\le a\varepsilon}\left|\sqrt{\varepsilon}\frac{m_1(T-t)\sigma^2(T-t,T)}{\sigma^2(T)}\int_0^T\frac{dB(s')}{m_1(s')}\right|\le Ca\varepsilon^{3/2}\left|\int_0^T\frac{dB(s')}{m_1(s')}\right|,\quad (2.98)$$

where C only depends on A and T. This term converges to 0 in the order $O(\varepsilon^{3/2})$.

To more rigorously demonstrate this, we use Martingale properties. If $g:(-\infty,T)\to\mathbf{R}$ is a continuous function we define $X(t)$, $t\ge 0$, by

$$X(t)=\int_{T-t}^T g(s)dB(s).\quad (2.99)$$

Then for any $\theta\in\mathbf{R}$,

$$X_\theta(t)=\exp\left[\theta X(t)-\frac{\theta^2}{2}\int_{T-t}^T ds\ g(s)^2\right]\text{ is an exponential Martingale}$$
$$\text{and } E[X_\theta(t)]=1.\quad (2.100)$$

According to the Markov inequality,

$$P(|X(T)|>M)=2P(X(T)>M)=2P\left(X_\theta(T)>\exp\left[\theta M-\frac{\theta^2}{2}\int_0^T g(s)^2ds\right]\right)$$
$$\le 2\exp\left[-\theta M+\frac{\theta^2}{2}\int_0^T g(s)^2ds\right],\text{ for }\theta>0.\quad (2.101)$$

The fact that $E[X_\theta(T)]=1$ is used in the inequality above. By choosing the

θ to minimize the right hand side[7] of the inequality (2.101), and substituting $g(s)$ by $1/m_1(s)$, we obtain

$$P\left(a\varepsilon^{3/2}\left|\int_0^T \frac{dB(s')}{m_1(s')}\right| > b\lambda\varepsilon/4\right) \leq 2\exp[-Cb^2\lambda^2/a^2\varepsilon], \tag{2.102}$$

where $C > 0$ is a constant depending only on A, T. We note here the choice of a, b needs to guarantee $b^2/a^2\varepsilon \to \infty$ as $\varepsilon \to 0$.

3. The third term is bounded as [8]

$$\begin{aligned}\sup_{0\leq t\leq a\varepsilon}\left|\sqrt{\varepsilon}\int_{T-t}^T\left(1-\frac{m_1(T-t)}{m_1(s')}\right)dB(s')\right| &\leq \sup_{0\leq t\leq a\varepsilon}\left|\sqrt{\varepsilon}\int_{T-t}^T\left(1-\frac{m_1(T)}{m_1(s')}\right)dB(s')\right| \\ &+ Ca\varepsilon^{3/2}\sup_{0\leq t\leq a\varepsilon}\left|\int_{T-t}^T\frac{dB(s')}{m_1(s')}\right|, \end{aligned} \tag{2.103}$$

where C depends only on A, T. To estimate the two terms on the right of (2.103), we define $X(t)$ and $X_\theta(t)$ the same as (2.99) and (2.100). However, instead of using Markov inequality, we use Doob's inequality here (See [25, page 13]). For any $\theta > 0$, we have

$$\begin{aligned}P\left(\sup_{0\leq t\leq t_0} X(t) > M\right) &\leq P\left(\sup_{0\leq t\leq t_0} X_\theta(t) > \exp\left[\theta M - \frac{\theta^2}{2}\int_{T-t_0}^T g(s)^2ds\right]\right) \\ &\leq \exp\left[-\theta M + \frac{\theta^2}{2}\int_{T-t_0}^T g(s)^2ds\right]. \end{aligned} \tag{2.104}$$

Optimizing the term on the right hand side of the inequality above with respect

[7] By choosing $\theta = M/\int_0^T g(s)^2ds$,

$$\exp\left[-\theta M + \frac{\theta^2}{2}\int_0^T g(s)^2ds\right] = \exp\left[-\frac{M^2}{2\int_0^T g(s)^2ds}\right].$$

[8] We note that the two terms on the right hand side of (2.103) are both martingales.

to θ we conclude that

$$P\left(\sup_{0\le t\le t_0}|X(t)|>M\right)\le 2\exp\left[-M^2\Big/2\int_{T-t_0}^{T}g(s)^2ds\right]. \tag{2.105}$$

Hence we have for the first term on the right hand side of (2.103) that

$$P\left(\sup_{0\le t\le a\varepsilon}\left|\sqrt{\varepsilon}\int_{T-t}^{T}\left(1-\frac{m_1(T)}{m_1(s')}\right)dB(s')\right|>b\lambda\varepsilon/4\right)\le 2\exp[-C_1b^2\lambda^2/a^3\varepsilon^2]\ ^{9}, \tag{2.106}$$

where $C_1>0$ is a constant depending only on A, T. Similarly we have

$$P\left(a\varepsilon^{3/2}\sup_{0\le t\le a\varepsilon}\left|\int_{T-t}^{T}\frac{dB(s')}{m_1(s')}\right|>b\lambda\varepsilon/4\right)\le 2\exp[-C_2b^2\lambda^2/a^3\varepsilon^2], \tag{2.107}$$

where the constant $C_2>0$ also depends only on A and T.

We now choose $a=\varepsilon^{-\alpha}, b=\varepsilon^{\beta}$ for some $\alpha,\beta>0$. Since $\mu>0$ it follows from (2.94) that the first term on the right hand side of (2.93) converges to $1-e^{-2\lambda\mu}$ as $\varepsilon\to 0$. We also see from the estimates of the previous paragraph that the second term on the right hand side of (2.93) converges to 0 as $\varepsilon\to 0$ provided $3\alpha+2\beta<2$ and $2\alpha+2\beta<1$. We have therefore shown that $\limsup_{\varepsilon\to 0}P(\inf_{0\le s\le T}Y_\varepsilon(s)>0)$ is bounded above by the right hand side of (2.84).

To obtain the corresponding lower bound we use the inequality

$$P\left(\inf_{0\le s\le T}Y_\varepsilon(s)>0\right)\ge P\left(\inf_{T-a\varepsilon\le s\le T}Y_\varepsilon(s)>0\right)-P\left(\inf_{0\le s\le T-a\varepsilon}Y_\varepsilon(s)<0\right)^{10}. \tag{2.108}$$

[9] Need to use the fact that $m_1(T)/m_1(s)-1<Ca\varepsilon$

[10] Since if $\inf_{T-a\varepsilon\le s\le T}Y_\varepsilon(s)>0$ then either $\inf_{0\le s\le T}Y_\varepsilon(s)>0$ or $\inf_{0\le s\le T-a\varepsilon}Y_\varepsilon(s)<0$.

Next, we use an inequality similar to (2.93):

$$P\left(\inf_{T-a\varepsilon\le s\le T} Y_\varepsilon(s) > 0\right) \ge P\left(\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t) > b\lambda\varepsilon\right) - P\left(\inf_{0\le t\le a\varepsilon} \tilde{Z}_\varepsilon(t) < -b\lambda\varepsilon\right)^{11}. \tag{2.109}$$

Similar to previously, we can choose $a = \varepsilon^{-\alpha}, b = \varepsilon^{\beta}$ with $3\alpha+2\beta < 2$ and $2\alpha+2\beta < 1$ to conclude that $\liminf_{\varepsilon\to 0} P\left(\inf_{T-a\varepsilon\le s\le T} Y_\varepsilon(s) > 0\right)$ is bounded below by the right hand side of (2.84).

Next, we would like to show that the second term on the right hand side (without the negative sign) of (2.108) vanishes as $\varepsilon \to 0$. Since $A(\cdot)$ is non-negative bounded, there is a positive constant C depending only on A, T such that the function $y(s)$ of (2.57) is bounded below by a linear function: $y(s) \ge C(T-s)y$ for $0 \le s \le T$. We can see this by observing that when $A(\cdot) \ge 0$,

$$\sigma^2(0,s) \ge m_2(0,s) \quad \text{and} \quad m_1(s,T)m_2(s,T) \ge \sigma^2(s,T),$$

thus in (2.57),

$$m_1(s,T)m_2(s,T)\sigma^2(0,s) - m_2(0,s)\sigma^2(s,T) \ge 0.$$

Therefore

$$y(s) \ge \frac{m_1(0,s)\sigma^2(s,T)}{\sigma^2(T)} y \ge C(T-s)y. \tag{2.110}$$

Since the expression of $Y_\varepsilon(s)$ in (2.87) has two stochastic components, the absolute value of one of them will have a large realization if $\inf_{0\le s\le T-\varepsilon^{1-\alpha}} Y_\varepsilon(s) < 0$. Hence

[11] Since if $\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t) > b\lambda\varepsilon$, then either $\inf_{T-a\varepsilon\le s\le T} Y_\varepsilon(s) > 0$ or $\inf_{0\le t\le a\varepsilon} \tilde{Z}_\varepsilon(t) < -b\lambda\varepsilon$.

there is a positive constant c depending only on A, T such that[12]

$$P\left(\inf_{0\le s\le T-\varepsilon^{1-\alpha}} Y_\varepsilon(s)<0\right)\le P\left(\left|\int_0^T \frac{dB(s')}{m_1(s')}\right|>\frac{cy}{\sqrt{\varepsilon}}\right) + P\left(\sup_{\varepsilon^{1-\alpha}\le t\le T}\left|\frac{1}{t}\int_{T-t}^T \frac{dB(s')}{m_1(s')}\right|>\frac{cy}{\sqrt{\varepsilon}}\right). \quad (2.111)$$

The first term on the right hand side can be bounded similarly to (2.102). For the second term, since the integral has t in its integral limit, which changes its variance, we split the interval $[\varepsilon^{1-\alpha}\le t\le T]$ into many small pieces. For any stochastic process $X(t)$,

$$P\left(\sup_{\varepsilon^{1-\alpha}\le t\le T}|X(t)|>cy/\sqrt{\varepsilon}\right)\le\sum_{k\ge1}P\left(\sup_{k\varepsilon^{1-\alpha}\le t\le(k+1)\varepsilon^{1-\alpha}}|X(t)|>cy/\sqrt{\varepsilon}\right). \quad (2.112)$$

By using Doob's inequality, similarly to (2.105), we see that for $k\ge1$,

$$P\left(\sup_{k\varepsilon^{1-\alpha}\le t\le(k+1)\varepsilon^{1-\alpha}}\left|\frac{1}{t}\int_{T-t}^T \frac{dB(s')}{m_1(s')}\right|>\frac{cy}{\sqrt{\varepsilon}}\right)\le 2\exp\left[-\frac{c_1ky^2}{\varepsilon^\alpha}\right] \quad (2.113)$$

where $c_1>0$ depends only on A, T. Therefore

$$P\left(\sup_{\varepsilon^{1-\alpha}\le t\le T}\left|\frac{1}{t}\int_{T-t}^T \frac{dB(s')}{m_1(s')}\right|>\frac{cy}{\sqrt{\varepsilon}}\right)\le\frac{2\exp\left[-c_1y^2/\varepsilon^\alpha\right]}{1-\exp\left[-c_1y^2/\varepsilon^\alpha\right]}, \quad (2.114)$$

which converges to 0 as $\varepsilon\to0$. Hence $\liminf_{\varepsilon\to0}P(\inf_{0\le s\le T}Y_\varepsilon(s)>0)$ is bounded below by the right hand side of (2.84). Hitherto, we have finished the proof of this proposition. □

Next we would like to derive an estimation on the ratio $G_{\varepsilon,D}(\lambda\varepsilon,y,0,T)/G_\varepsilon(\lambda\varepsilon,y,0,T)$ with $\varepsilon\ne0$ which is uniform as $\lambda\to0$.

Lemma II.20. *Assume the function $A(\cdot)$ is non-negative and that $0<\lambda\le1$, $0<$*

[12]Notice that $m_1(s),\sigma^2(T)$ behaves like constant, $\sigma^2(s,T)$ like $(T-s)$. Also we have substituted $a=\varepsilon^{-\alpha}$ into the equation.

$\varepsilon \le T$, $y > 0$. *Let* $\Gamma : \mathbf{R}^+ \times \mathbf{R}^+ \to \mathbf{R}^+$ *be the function* $\Gamma(a,b) = 1$ *if* $b > a^{-1/4}$ *and otherwise* $\Gamma(a,b) = a^{1/8}$. *Then there is a constant* C *depending only on* AT *such that*

$$\frac{G_{\varepsilon,D}(\lambda\varepsilon, y, 0, T)}{G_{\varepsilon}(\lambda\varepsilon, y, 0, T)} \le \\ 1 - \exp\left[-2\lambda\left\{1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right\}\right] + C\lambda\Gamma\left(\frac{\varepsilon}{T}, \frac{y}{T}\right)\left[1 + \frac{y}{T}\right]. \quad (2.115)$$

Remark II.21. *Intuitively, in (2.115), the first part on the right hand side is the same as the limit in (2.84). The second part mainly depends on* ε *and* y*, since larger values of* ε *and* y *will pull the trajectory of* $Y_\varepsilon(s)$ *away from the boundary 0, thus leading to a higher probability for* $\inf_{0\le s\le T} Y_\varepsilon(s) > 0$.

Proof of Lemma II.20. The process $Y_\varepsilon(s)$ we are considering is given in equation (2.87). We have observed from Proposition II.18, the derivative of $y(s)$ towards the end $s = T$ determines the main behavior of $P(\inf_{0\le s\le T} Y_\varepsilon(s) > 0)$. It will be convenient to make a change of variable so that we work on a process with time 0 being the point we mainly focus on. Also, we would like to express $\int_s^T \frac{dB(s')}{m_1(s')}$ as a Brownian motion. In order to serve this purpose, we make a change of variable and define a variable t such that $s(t)$ satisfies

$$\frac{ds}{dt} = -\left[\frac{m_1(s)}{m_1(T)}\right]^2, \quad s(0) = T. \quad (2.116)$$

We note that when $t = 0$, $s = T$. also $s \simeq T - t$ if t is small in the sense that $\lim_{t\to 0}(T - s)/t = 1$ [13]. By solving the differential equation (2.116), we can obtain a

[13]It is obvious that $s(t)$ is a decreasing function of t. Since $ds/dt = -\exp\left(-\int_s^T 2A(s')ds'\right) > -1$ and $-\exp\left(-\int_s^T 2A(s')ds'\right) < -\exp\left(-2A(T-s)\right) < -(1 - 2A(T-s))$, therefore $1 - 2A(T-s) < d(T-s)/dt < 1$, which implies $(1 - e^{-2At})/2A < T - s < t$.

time point $\tilde{T} = \sigma^2(0,T)$ for the variable t such that $s(\tilde{T}) = 0$. Also, we have

$$m_1(T)\int_s^T \frac{dB(s')}{m_1(s')} = \int_0^t d\tilde{B}(t')^{14}, \tag{2.117}$$

where $\tilde{B}(\cdot)$ is a Brownian motion. Therefore from (2.86) we can define $\tilde{Y}_\varepsilon(t) = Y_\varepsilon(s(t))$,

$$\tilde{Y}_\varepsilon(t) = \tilde{y}(t) + \sqrt{\varepsilon}\left[\frac{m_1(s)\sigma^2(s,T)}{m_1(T)\sigma^2(T)}\int_0^{\tilde{T}} d\tilde{B}(t') - \frac{m_1(s)}{m_1(T)}\int_0^t d\tilde{B}(t')\right],^{15} \tag{2.118}$$

where $\tilde{y}(t) = y(s(t))$. From now on, we basically work on $\tilde{Y}_\varepsilon(t)$.

Since $\tilde{Y}_\varepsilon(t)$ tends to exit through the boundary 0 in a time of order $O(\varepsilon)$, we consider any a for which $0 < a\varepsilon \le \tilde{T}$ and observe that

$$P\left(\inf_{0\le s\le T} Y_\varepsilon(s) > 0\right) \le P\left(\inf_{0\le t\le a\varepsilon} \tilde{Y}_\varepsilon(t) > 0\right). \tag{2.119}$$

The magnitude of $\tilde{Y}_\varepsilon(t)$ depends on both the deterministic part $\tilde{y}(t)$ and the stochastic part. For different realization of the stochastic part, the estimation on the probability above is different. For any $M > 0$,

$$\begin{aligned} P\left(\inf_{0\le t\le a\varepsilon} \tilde{Y}_\varepsilon(t) > 0\right) = {} & P\left(\inf_{0\le t\le a\varepsilon} \tilde{Y}_\varepsilon(t) > 0;\ \sup_{0\le t\le a\varepsilon}\left|\int_0^t d\tilde{B}(t')\right| \le M\right) \\ & + P\left(\inf_{0\le t\le a\varepsilon} \tilde{Y}_\varepsilon(t) > 0;\ \sup_{0\le t\le a\varepsilon}\left|\int_0^t d\tilde{B}(t')\right| > M\right). \end{aligned} \tag{2.120}$$

For the first term on the right hand side of (2.120), we compare $\tilde{Y}_\varepsilon(t)$ with a process $\tilde{Y}_{0,\varepsilon}(t)$ which is defined as

$$\tilde{Y}_{0,\varepsilon}(t) = \tilde{y}(t) + \frac{C\sqrt{\varepsilon}Mt}{T} + \sqrt{\varepsilon}\frac{m_1(s)\sigma^2(s,T)}{m_1(T)\sigma^2(T)}\int_{a\varepsilon}^{\tilde{T}} d\tilde{B}(t') - \sqrt{\varepsilon}\frac{m_1(s)}{m_1(T)}\int_0^t d\tilde{B}(t'), \tag{2.121}$$

[14] $d\tilde{B}(t)$ can be defined as $d\tilde{B}(t) = m_1(T)dB(s)/m_1(s)$.

[15] In this equation and below, s can be seen as $s(t)$.

where C depends only on AT. To derive C, we have

$$\sqrt{\varepsilon}\frac{m_1(s)\sigma^2(s,T)}{m_1(T)\sigma^2(T)}\int_0^{a\varepsilon} d\tilde{B}(t') < \sqrt{\varepsilon}M\frac{\int_s^T m_1^2(s',T)ds'}{\int_0^T m_1^2(s',T)ds'} < \sqrt{\varepsilon}M\frac{t\cdot\exp[2AT]}{T}.$$

The C in (2.121) can be chosen as $\exp[2AT]$. It is easy to see that $\tilde{Y}_{0,\varepsilon}(t) \geq \tilde{Y}_\varepsilon(t)$ for $0 \leq t \leq a\varepsilon$ under the condition that $\sup_{0\leq t\leq a\varepsilon}\left|\int_0^t d\tilde{B}(t')\right| \leq M$. Therefore,

$$P\left(\inf_{0\leq t\leq a\varepsilon}\tilde{Y}_\varepsilon(t) > 0;\ \sup_{0\leq t\leq a\varepsilon}\left|\int_0^t d\tilde{B}(t')\right| \leq M\right) \leq P\left(\inf_{0\leq t\leq a\varepsilon}\tilde{Y}_{0,\varepsilon}(t) > 0\right) \tag{2.122}$$

We would like to define a process $Z_\varepsilon(t)$ as in (2.91),

$$dZ_\varepsilon(t) = \mu dt - \sqrt{\varepsilon}d\tilde{B}(t), \quad Z_\varepsilon(0) = \lambda\varepsilon[1 + Ca\varepsilon/T], \tag{2.123}$$

with $\mu = \mu_{rand}$ a well-chosen constant drift such that $m_1(s)Z_\varepsilon(t)/m_1(T) \geq \tilde{Y}_{0,\varepsilon}(t)$ over $0 \leq t \leq a\varepsilon$ [16]. To do this, we need to estimate the derivative of $\tilde{y}(t)$ over the interval $[0, a\varepsilon]$. At $t = 0$, $d\tilde{y}(t)/dt = O(\varepsilon) + 1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}$; for $0 \leq t \leq a\varepsilon$, according to equation (2.88),

$$\frac{d\tilde{y}}{dt} < 1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)} + \frac{Ca\varepsilon}{T}\left[1 + \frac{y}{T}\right] + \frac{C\lambda\varepsilon}{T}{}^{[17]}. \tag{2.124}$$

Therefore, we can choose μ_{rand} as

$$\mu_{rand} = 1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)} + \frac{Ca\varepsilon}{T}\left[1 + \frac{y}{T}\right] + \frac{C\lambda\varepsilon}{T} + \frac{C\sqrt{\varepsilon}}{T}\left[M + \left|\int_{a\varepsilon}^{\tilde{T}} d\tilde{B}(t')\right|\right], \tag{2.125}$$

[16] Since $m_1(T)/m_1(s) \leq 1 + Ca\varepsilon/T$, we set $Z_\varepsilon(0) = \lambda\varepsilon[1 + Ca\varepsilon/T]$.

[17] $\left|\lambda\varepsilon \cdot m_1(s,T)(1 + A(s)\sigma^2(0,s))/\sigma^2(T)\right| \leq \lambda\varepsilon\exp(AT)\cdot(1 + A\cdot C_1T)/C_2T \leq C\lambda\varepsilon/T$, where C_1, C_2 and C are constants depending only on AT. $\left|y\cdot m_1(0,s)A(s)\sigma^2(s,T)/\sigma^2(T)\right| \leq Cy\exp(AT)Aa\varepsilon/T \leq Cya\varepsilon/T^2$, where C is a constant depending only on AT. $\left|m_1(s,T)m_2(s,T)(1 + A(s)\sigma^2(0,s))/\sigma^2(T)\right| \leq Ca\varepsilon/T$ for C being a constant depending only on AT. $\left|-(1 + A(s)m_2(0,s))\sigma^2(s,T)/\sigma^2(T)\right| \leq Ca\varepsilon/T$ for C being a constant depending only on AT.

in which C is a constant only depending on AT such that $m_1(s)Z_\varepsilon(t)/m_1(T) \geq \tilde{Y}_{0,\varepsilon}(t)$[18]. Therefore,

$$P\left(\inf_{0\leq t\leq a\varepsilon} \tilde{Y}_{0,\varepsilon}(t) > 0\right) = P\left(\inf_{0\leq t\leq a\varepsilon} \frac{m_1(T)}{m_1(s)}\tilde{Y}_{0,\varepsilon}(t) > 0\right)$$
$$\leq E\left[P\left(\inf_{0\leq t\leq a\varepsilon} Z_\varepsilon(t) > 0\middle|\mu = \mu_{rand}, Z_\varepsilon(0) = \lambda\varepsilon[1 + Ca\varepsilon/T]\right)\right]. \quad (2.126)$$

To bound the right hand side of (2.126), we use an identity

$$P\left(\inf_{0\leq t\leq a'\varepsilon} Z_\varepsilon(t) > 0\middle|Z_\varepsilon(0) = \lambda'\varepsilon\right) =$$
$$\left\{1 - e^{-2\mu\lambda'}\right\}\frac{1}{\sqrt{2\pi}}\int_{[\lambda'-\mu a']/\sqrt{a'}}^{\infty} e^{-z^2/2}dz + \frac{1}{\sqrt{2\pi}}\int_{[-\lambda'-\mu a']/\sqrt{a'}}^{[\lambda'-\mu a']/\sqrt{a'}} e^{-z^2/2}dz. \quad (2.127)$$

From this identity, we can obtain an upper bound

$$P\left(\inf_{0\leq t\leq a'\varepsilon} Z_\varepsilon(t) > 0\middle|Z_\varepsilon(0) = \lambda'\varepsilon\right) \leq 1 - e^{-2\mu\lambda'} + \frac{2\lambda'}{\sqrt{2\pi a'}}. \quad (2.128)$$

This bound plays an important role in deriving the last term on the right hand side of (2.115). Using (2.128), we estimate the right hand side of (2.126) when $a = \min\left[(T/\varepsilon)^\alpha, \tilde{T}/\varepsilon\right]$ for some α satisfying $0 < \alpha < 1$ [19]. In that case $\lambda' = \lambda[1 + Ca\varepsilon/T] \leq \lambda[1 + C]$ for some constant C depending only on AT. Taking $M = C_1\sqrt{T}$ in (2.125) where C_1 is a constant depending only on AT, then we have for $0 < \lambda \leq 1$,

[18] Note that

$$\frac{m_1(T)}{m_1(s)}\left(1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right) \leq (1 + Ca\varepsilon/T)\cdot\left(1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right)$$
$$\leq \left(1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right) + \frac{Ca\varepsilon}{T}\left[1 + \frac{y}{T}\right].$$

[19] With a defined in this way, $a\varepsilon \to 0$ as $\varepsilon \to 0$, but $a\varepsilon$ converges to 0 slower than ε.

$0 < \varepsilon \le T$,

$$
\begin{aligned}
&P\left(\inf_{0\le t\le a\varepsilon} \tilde{Y}_{0,\varepsilon}(t) > 0\right) \le E\left[1 - e^{-2\mu_{rand}\lambda(1+Ca\varepsilon/T)} + \frac{2\lambda(1+C)}{\sqrt{2\pi a}}\right] \\
&\le 1 - \exp\left[-2\lambda\left\{1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right\}\right] + C_2\lambda\left[\left(\frac{\varepsilon}{T}\right)^{1-\alpha}\left(1+\frac{y}{T}\right) + \left(\frac{\varepsilon}{T}\right)^{1/2} + \left(\frac{\varepsilon}{T}\right)^{\alpha/2}\right] \\
&\le 1 - \exp\left[-2\lambda\left\{1 - \frac{m_2(T)}{\sigma^2(T)} + \frac{m_1(T)y}{\sigma^2(T)}\right\}\right] + C_3\lambda\left[\left(\frac{\varepsilon}{T}\right)^{1-\alpha}\left(1+\frac{y}{T}\right) + \left(\frac{\varepsilon}{T}\right)^{\alpha/2}\right].
\end{aligned}
\tag{2.129}
$$

In this inequality, $\left(\frac{\varepsilon}{T}\right)^{1-\alpha}\left(1+\frac{y}{T}\right)+\left(\frac{\varepsilon}{T}\right)^{1/2}$ comes from $e^{-2\mu_{rand}\lambda(1+Ca\varepsilon/T)}$, $\left(\frac{\varepsilon}{T}\right)^{\alpha/2}$ comes from $\frac{2\lambda(1+C)}{\sqrt{2\pi a}}$, and the third inequality is true since $(\varepsilon/T)^{1/2}$ can be absorbed by $(\varepsilon/T)^{\alpha/2}$ [20]. So far, we have finished the estimation of the first term on the right hand side of (2.120).

Next we estimate the second term. To do this, we introduce the stopping time τ defined by

$$
\tau = \inf\left\{t < \tilde{T} : \left|\int_0^t d\tilde{B}(t')\right| > M\right\}. \tag{2.130}
$$

Hence the second term is bounded above by $P\left(\inf_{0\le t\le \tau} \tilde{Y}_\varepsilon(t) > 0; \tau < a\varepsilon\right)$. To estimate the probability, we usually need to compare the process with other ones of constant drifts. From (2.118), we see that the random drift of $\tilde{Y}_\varepsilon(t)$ is impacted by the realization of the integral $\int_0^{\tilde{T}} d\tilde{B}(t')$. By using this as an indicator, we compare $\tilde{Y}_\varepsilon(t)$ with different simpler stochastic processes. Intuitively, the larger $\int_0^{\tilde{T}} d\tilde{B}(t')$ is, the larger the probability for the compared process to be positive, but at the same

[20] $(\varepsilon/T)^{1/2} = o(\varepsilon/T)^{\alpha/2}$ as $\varepsilon \to 0$.

time, the probability for a large realization of $\int_0^{\tilde{T}} d\tilde{B}(t')$ is small.

$$\begin{aligned}
& P\left(\inf_{0\le t\le\tau} \tilde{Y}_\varepsilon(t) > 0; \tau < a\varepsilon\right) \\
&= \sum_{n=1}^{\infty} P\left(\inf_{0\le t\le\tau} \tilde{Y}_\varepsilon(t) > 0; \tau < a\varepsilon, (n-1)M_1 \le \sup_{\tau\le t\le\tau+\tilde{T}} \left|\int_\tau^t d\tilde{B}(t')\right| < nM_1\right) \\
&\le \sum_{n=1}^{\infty} P\left(\inf_{0\le t\le\tau} \tilde{Y}_{n,\varepsilon}(t) > 0; \tau < a\varepsilon\right) P\left((n-1)M_1 \le \sup_{\tau\le t\le\tau+\tilde{T}} \left|\int_\tau^t d\tilde{B}(t')\right| < nM_1\right) \\
&= \sum_{n=1}^{\infty} P\left(\inf_{0\le t\le\tau} \tilde{Y}_{n,\varepsilon}(t) > 0; \tau < a\varepsilon\right) P\left((n-1)M_1 \le \sup_{0\le t\le\tilde{T}} \left|\int_0^t d\tilde{B}(t')\right| < nM_1\right),
\end{aligned} \tag{2.131}$$

where $\tilde{Y}_{n,\varepsilon}$ is given by the formula

$$\tilde{Y}_{n,\varepsilon}(t) = \tilde{y}(t) + \frac{C\sqrt{\varepsilon}(M+nM_1)t}{T} - \sqrt{\varepsilon}\frac{m_1(s)}{m_1(T)}\int_0^t d\tilde{B}(t'), \tag{2.132}$$

and the constant C depends only on AT. Here, we have used the strong Markov property of τ which implies that $\left\{\tilde{B}(t) : 0 < t \le \tau\right\}$ are independent of the variable $\sup_{\tau\le t\le\tau+\tilde{T}}\left|\int_0^t d\tilde{B}(t')\right|$. At the hitting time τ, we have

$$\tilde{Y}_{n,\varepsilon}(\tau) = \tilde{y}(\tau) + \frac{C\sqrt{\varepsilon}(M+nM_1)\tau}{T} \pm M\sqrt{\varepsilon}\frac{m_1(s(\tau))}{m_1(T)}. \tag{2.133}$$

We choose $M_1 = \sqrt{T}$ and $M = C_1\sqrt{T}$ where C_1 is a constant depending only on AT. We also notice that there exists constant C_2 depending only on AT, such that $\tilde{y}(\tau) \le C_2(1 + y/T)\tau$. Therefore, we can find a lower bound of τ for different values of n, τ_n, such that the hitting time τ has to be larger than τ_n in order for $\tilde{Y}_{n,\varepsilon}(\tau)$ to be non-negative when (2.133) holds with *minus* sign,

$$\tau > \tau_n = cT\sqrt{\frac{\varepsilon}{T}}\left[1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right]^{-1}, \tag{2.134}$$

where c is a constant depending only on AT. We notice that $1+y/T$ and n influence the magnitude of τ_n. If $\tau_n > a\varepsilon$, then the probability for $\inf_{0\le t\le\tau} \tilde{Y}_{n,\varepsilon}(t) < 0$ for $\tau < a\varepsilon$ is 0. We observe now that if $\alpha < 1/2$ then $\tau_n > a\varepsilon$ provided

$$1+\frac{y}{T}+n\sqrt{\frac{\varepsilon}{T}} \le 2c_1\left(\frac{T}{\varepsilon}\right)^{1/2-\alpha} \quad \text{for } c_1 = c/2 > 0 \text{ depending only on } AT. \tag{2.135}$$

If $\tau_n > a\varepsilon$, i.e., the inequality in (2.135) happens, then $P\left(\inf_{0\le t\le\tau} \tilde{Y}_{n,\varepsilon}(t) > 0; \tau < a\varepsilon\right)$ as in (2.131) equals to 0. Therefore, we only need to concern ourself with the scenario that $1+\frac{y}{T}+n\sqrt{\frac{\varepsilon}{T}} > 2c_1\left(\frac{T}{\varepsilon}\right)^{1/2-\alpha}$. We note that (2.133) can hold with $\tau < a\varepsilon$ and a "$-$" in front of the last term only if (necessary but not sufficient)

$$1+\frac{y}{T} \ge c_1\left(\frac{T}{\varepsilon}\right)^{1/2-\alpha} \text{ or } n \ge c_1\left(\frac{T}{\varepsilon}\right)^{1-\alpha}, \tag{2.136}$$

which corresponds to the union of the regions (I), (II), (IV) in Figure 2.1.

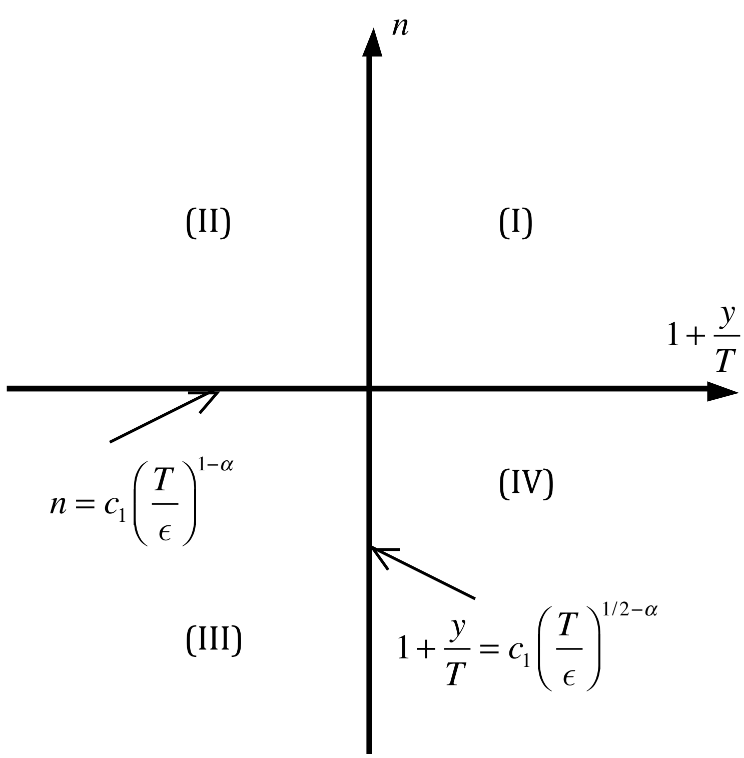


Figure 2.1: Four possible combinations of $1 + y/T$ and n.

In the case when $\tau_n < a\varepsilon$ we see from (2.133) that there is a constant C depending only on AT and

$$P\left(\inf_{0\le t\le\tau_n} \tilde{Y}_{n,\varepsilon}(t) > 0\right) \le P\left(\inf_{0\le t\le\tau_n} Z_\varepsilon(t) > 0 \middle| \mu = \mu_n, Z_\varepsilon(0) = \lambda\varepsilon[1 + Ca\varepsilon/T]\right), \tag{2.137}$$

where $Z_\varepsilon(\cdot)$ is the solution to the SDE in (2.91). The drift μ_n is given by the formula

$$\mu_n = C\left[1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right] \text{ where } C \text{ depends only on } AT. \tag{2.138}$$

Following (2.128), (2.137) and (2.138), we have

$$P\left(\inf_{0\le t\le\tau_n} \tilde{Y}_{n,\varepsilon}(t) > 0\right) \le C_1\lambda\left[1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}} + \left(\frac{\varepsilon}{\tau_n}\right)^{1/2}\right] \le C_2\lambda\left[1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right], \tag{2.139}$$

where constants C_1, C_2 only depends on AT. The second inequality holds since

$$\left(\frac{\varepsilon}{\tau_n}\right)^{1/2} = C\left[\sqrt{\frac{\varepsilon}{T}}\left(1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right)\right]^{1/2} \le C'\left(1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right), \tag{2.140}$$

in which C, C' are constants depending only on AT.

Further, for large n values, i.e., $n \ge c_1(T/\varepsilon)^{1-\alpha}$ (which corresponds to region (I) and (II) in Figure 2.1), we conclude from (2.139) that

$$\begin{aligned}
&\sum_{n\ge c_1(T/\varepsilon)^{1-\alpha}} P\left(\inf_{0\le t\le\tau_n} \tilde{Y}_{n,\varepsilon}(t) > 0\right) P\left((n-1)M_1 \le \left|\sup_{0\le t\le\tilde{T}} \int_0^t d\tilde{B}(t')\right| < nM_1\right) \\
&\le C\lambda \sum_{n\ge c_1(T/\varepsilon)^{1-\alpha}} \left[1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right] e^{-n^2/2} \le C_1\lambda\left(1 + \frac{y}{T}\right)\exp\left[-c_1\left(\frac{T}{\varepsilon}\right)^{2(1-\alpha)}\right],
\end{aligned} \tag{2.141}$$

where the constants C_1, c_1 depend only on AT. When n is small, i.e., $n < c_1\left(\frac{T}{\varepsilon}\right)^{1-\alpha}$, and $1+y/T > C_1(T/\varepsilon)^{1/2-\alpha}$, we have $n\sqrt{\varepsilon/T} \leq 1+y/T$, then according to (2.139),

$$P\left(\inf_{0\leq t\leq \tau_n} \tilde{Y}_{n,\varepsilon}(t) > 0\right) \leq C_3\lambda\left[1+\frac{y}{T}\right], \text{ for } n < c_1(T/\varepsilon)^{1-\alpha}. \tag{2.142}$$

Therefore,

$$\begin{aligned}
&\sum_{n<c_1(T/\varepsilon)^{1-\alpha}} P\left(\inf_{0\leq t\leq \tau_n} \tilde{Y}_{n,\varepsilon}(t) > 0\right) P\left((n-1)M_1 \leq \left|\sup_{0\leq t\leq \tilde{T}} \int_0^t d\tilde{B}(t')\right| < nM_1\right) \\
&\leq P\left(\inf_{0\leq t\leq \tau_N} \tilde{Y}_{N,\varepsilon}(t) > 0\right) \sum_{n<c_1(T/\varepsilon)^{1-\alpha}} P\left((n-1)M_1 \leq \left|\sup_{0\leq t\leq \tilde{T}} \int_0^t d\tilde{B}(t')\right| < nM_1\right) \\
&\leq P\left(\inf_{0\leq t\leq \tau_N} \tilde{Y}_{N,\varepsilon}(t) > 0\right) \leq C_3\lambda\left[1+\frac{y}{T}\right],
\end{aligned} \tag{2.143}$$

where $N = \lfloor c_1\left(\frac{T}{\varepsilon}\right)^{1-\alpha}\rfloor$.

Next we consider the situation when (2.133) holds with the *plus* sign. One sees that

$$\begin{aligned}
&P\left(\inf_{0\leq t\leq \tau} \tilde{Y}_{n,\varepsilon}(t) > 0; \tau < a\varepsilon, \int_0^\tau d\tilde{B}(t') = -M\right) \\
&\quad \leq P\left(\inf_{0\leq t\leq \tau} Z_\varepsilon(t) > 0; \tau < a\varepsilon, Z_\varepsilon(\tau) \geq M\sqrt{\varepsilon}\middle|\mu = \mu_n, Z_\varepsilon(0) = \lambda\varepsilon[1+Ca\varepsilon/T],\right)
\end{aligned} \tag{2.144}$$

where μ_n is given by (2.138). Here, $Z_\varepsilon(t) \approx m_1(T)\tilde{Y}_{n,\varepsilon}(t)/m_1(s(t))$. By observing the right hand side of (2.144), we see that it is bounded by the probability that the diffusion $Z_\varepsilon(\cdot)$ started at $\lambda\varepsilon[1+O(\varepsilon^{1-\alpha})]$ exits the interval $[0, C_1T(\varepsilon/T)^{1/2}]$ through the boundary $C_1T(\varepsilon/T)^{1/2}$ in time less than $T(\varepsilon/T)^{1-\alpha}$. When $\alpha < 1/2$, $T(\varepsilon/T)^{1-\alpha}$ converges to 0 faster than $C_1T(\varepsilon/T)^{1/2}$, thus we are essentially estimating the probability for a diffusion to exit a large interval in a very short period of time, which converges to 0 as $\varepsilon/T \to 0$. To find an expression for (2.144), we first choose C_1 large enough so that $Z_\varepsilon(0) < C_1 T(\varepsilon/T)^{1/2}/2$ for any ε satisfying $0 < \varepsilon \le T$. The process associated to the probability we are studying is composed of two parts: 1) exiting through $C_1 T(\varepsilon/T)^{1/2}/2$ instead of 0; 2) moving up to $C_1 T(\varepsilon/T)^{1/2}$ from $C_1 T(\varepsilon/T)^{1/2}/2$ within a period of time $a\varepsilon$.

$$
\begin{aligned}
P\left(\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t) > 0, \sup_{0\le t\le a\varepsilon} Z_\varepsilon(t) \ge \Lambda'\varepsilon | Z_\varepsilon(0) = \lambda'\varepsilon\right) \le \\
P(0 < Z_\varepsilon(t) < \Lambda'\varepsilon/2, t < \tau, Z_\varepsilon(\tau) = \Lambda'\varepsilon/2 | Z_\varepsilon(0) = \lambda'\varepsilon) \cdot \\
P\left(\sup_{0\le t\le a\varepsilon} Z_\varepsilon(t) \ge \Lambda'\varepsilon \middle| Z_\varepsilon(0) = \Lambda'\varepsilon/2\right). \quad (2.145)
\end{aligned}
$$

We now compute these two probabilities individually.

For part 1), it is easy to see that for $0 < \lambda' < \Lambda'$,

$$
P(0 < Z_\varepsilon(t) < \Lambda'\varepsilon, t < \tau, Z_\varepsilon(\tau) = \Lambda'\varepsilon \big| Z_\varepsilon(0) = \lambda'\varepsilon) = \frac{1 - e^{-2\mu\lambda'}}{1 - e^{-2\mu\Lambda'}}. \quad (2.146)
$$

By plugging in $\Lambda' = C_1(T/\varepsilon)^{1/2}, \mu = \mu_n$ and $\lambda' = \lambda[1 + Ca\varepsilon/T]$, whence $\mu\Lambda' \ge c$ for some positive c depending only on AT [21], we conclude that

$$
P(0 < Z_\varepsilon(t) < \Lambda'\varepsilon/2, t < \tau, Z_\varepsilon(\tau) = \Lambda'\varepsilon/2 | Z_\varepsilon(0) = \lambda'\varepsilon) \le C\lambda\left[1 + \frac{y}{T} + n\sqrt{\frac{\varepsilon}{T}}\right] \quad (2.147)
$$

for some constant C depending only on AT.

For part 2),

$$
P\left(\sup_{0\le t\le a\varepsilon} Z_\varepsilon(t) \ge \Lambda'\varepsilon \middle| Z_\varepsilon(0) = \Lambda'\varepsilon/2\right) \le C\exp\left[-\frac{\Lambda'^2}{32a}\right] \quad (2.148)
$$

for some universal constant C provided $\mu a < \Lambda'/4$ [22]. We observe that $\mu_n a < \Lambda'/4$ is

[21] Thus $1 - e^{-2\mu\Lambda'} > 1 - e^{-2c}$.

[22] The method of image can be applied here, and since since $\mu a < \Lambda'/4$

$$
P\left(\sup_{0\le t\le a\varepsilon} Z_\varepsilon(t) \ge \Lambda'\varepsilon \middle| Z_\varepsilon(0) = \Lambda'\varepsilon/2\right) \le C\exp\left[-\frac{(\Lambda'\varepsilon - \Lambda'\varepsilon/2 - \mu a\varepsilon)^2}{2\varepsilon\cdot a\varepsilon}\right] \le C\exp\left[-\frac{(\Lambda'\varepsilon/4)^2}{2a}\right].
$$

naturally implied if (2.135) holds. We conclude that if (2.135) holds, then

$$P\left(\inf_{0\le t\le a\varepsilon} Z_\varepsilon(t)>0,\ \sup_{0\le t\le a\varepsilon} Z_\varepsilon(t)\ge \Lambda'\varepsilon \middle| Z_\varepsilon(0)=\lambda'\varepsilon\right) \\ \le C_2\lambda\left[1+\frac{y}{T}+n\sqrt{\frac{\varepsilon}{T}}\right]\exp\left[-c_2\left(\frac{T}{\varepsilon}\right)^{1-\alpha}\right]. \quad (2.149)$$

If (2.135) does not hold, we can argue as before to obtain an inequality similar to (2.141). On choosing $\alpha = 1/4$, (2.115) follows from (2.129), (2.141) and (2.149). $\square$

Lemma II.20 demonstrates an upper bound for the ratio of the Dirichlet Green's function to the full space Green's function. Next we try to derive a lower bound for this ratio.

Lemma II.22. *Assume the function $A(\cdot)$ is non-negative and that $0<\lambda\le 1$, $0<\varepsilon\le T$, $y>0$. Then there are positive constants C, c depending only on AT such that if $\gamma = c(T/\varepsilon)^{1/8}(y/T)\ge 5$ then*

$$\frac{G_{\varepsilon,D}(\lambda\varepsilon,y,0,T)}{G_\varepsilon(\lambda\varepsilon,y,0,T)} \ge [1+e^{-\gamma^2/4}]^{-2}\left(1-\exp\left[-\frac{2\lambda}{1+C(\varepsilon/T)^{1/8}}\left\{1-\frac{m_2(T)}{\sigma^2(T)}+\frac{m_1(T)y}{\sigma^2(T)}\right\}\right]\right). \quad (2.150)$$

Proof. We choose $a=\min\left[(T/\varepsilon)^\alpha, \tilde{T}/\varepsilon\right]$ with $0<\alpha<1$ as in Lemma II.20. Also, we observe that there is a constant $c>0$ depending only on AT such that $\tilde{y}(t)\ge cty/T$ for $0\le t\le \tilde{T}$. Intuitively, as long as the magnitude of the Brownian motion is small enough, $\tilde{Y}_\varepsilon(\cdot)$ can be kept away from the boundary 0. More specifically, there exists a constant $c_1>0$ depending only on AT such that the process $\tilde{Y}_\varepsilon(\cdot)$ of (2.118) satisfies $\tilde{Y}_\varepsilon(t)>0$ for $a\varepsilon\le t\le \tilde{T}$ if $\mathcal{E}$ holds, where the event $\mathcal{E}$ is defined as

$$\left|\int_0^{a\varepsilon} d\tilde{B}(t')\right| < c_1\sqrt{T}\left(\frac{\varepsilon}{T}\right)^{1/2-\alpha}\frac{y}{T} \text{ and} \\ \sup_{a\varepsilon\le t\le (k+1)a\varepsilon}\left|\int_{a\varepsilon}^t d\tilde{B}(t')\right| \le c_1 k\sqrt{T}\left(\frac{\varepsilon}{T}\right)^{1/2-\alpha}\frac{y}{T}, \text{ for } k=1,2,... \quad (2.151)$$

Actually, this can be shown as follows: for $a\varepsilon \le t \le \tilde{T}$,

$$\tilde{Y}_\varepsilon(t) \ge \frac{cty}{T} - C\frac{t}{T}\sqrt{\varepsilon}c_1 \left\lceil \frac{\tilde{T}}{a\varepsilon} \right\rceil \sqrt{T}\left(\frac{\varepsilon}{T}\right)^{1/2-\alpha}\frac{y}{T} - C\sqrt{\varepsilon}\left\lceil \frac{t}{a\varepsilon} \right\rceil c_1\sqrt{T}\left(\frac{\varepsilon}{T}\right)^{1/2-\alpha}\frac{y}{T}$$
$$= \frac{cty}{T} - Cc_1\frac{ty}{T} - Cc_1\frac{ty}{T},$$

where $\lceil x \rceil := \min_{y\in\mathbb{N}}\{y \ge x\}$, and C is a constant depending only on AT. Thus as long as c_1 is small enough, $\mathcal{E}$ guarantees that $\tilde{Y}_\varepsilon(t) > 0$ for $a\varepsilon \le t \le \tilde{T}$.

It follows that

$$P\left(\inf_{0\le s\le T} Y_\varepsilon(s) > 0\right) = P\left(\inf_{0\le t\le \tilde{T}} \tilde{Y}_\varepsilon(t) > 0\right) \ge P\left(\inf_{0\le t\le a\varepsilon} \tilde{Y}_\varepsilon(t) > 0; \mathcal{E}\right). \quad (2.152)$$

It is easy to see from (2.118) that on the event $\mathcal{E}$ there exists a constant $C' > 0$ depending only on AT such that

$$\tilde{Y}_\varepsilon(t) > 0 \text{ if } \tilde{Z}_\varepsilon(t) = \frac{\tilde{y}(t)}{1 + C'a\varepsilon/T} - C'c_1 t\frac{y}{T} - \sqrt{\varepsilon}\int_0^t d\tilde{B}(t') > 0, \text{ for } 0 < t \le a\varepsilon, \quad (2.153)$$

where c_1 is the constant of (2.151). Here $\tilde{Z}_\varepsilon(t) \le m_1(T)\tilde{Y}_\varepsilon(t)/m_1(s)$. We conclude from (2.151), (2.153) that

$$P\left(\inf_{0\le s\le T} Y_\varepsilon(s) > 0\right)$$
$$\ge P\left(\tilde{Z}_\varepsilon(t) > 0, 0 < t \le a\varepsilon; \left|\int_0^{a\varepsilon} d\tilde{B}(t')\right| < c_1\sqrt{T}\left(\frac{\varepsilon}{T}\right)^{1/2-\alpha}\frac{y}{T}\right) P(\mathcal{E}). \quad (2.154)$$

We will bound the two terms on the right hand side of the inequality above separately.

In order to bound $P(\mathcal{E})$ from below, we consider for $\gamma > 0$ the event $\mathcal{E}_\gamma$ defined by

$$\left|\int_0^1 d\tilde{B}(t')\right| < \gamma \text{ and } \sup_{1\le t\le (k+1)}\left|\int_1^t d\tilde{B}(t')\right| < k\gamma \quad \text{for } k = 1, 2, ... \quad (2.155)$$

Then we have

$$P(\mathcal{E}) = P(\mathcal{E}_\gamma) \quad \text{where} \quad \gamma = c_1 \left(\frac{T}{\varepsilon}\right)^{\alpha/2} \left(\frac{y}{T}\right). \tag{2.156}$$

According to Doob's inequality,

$$P\left(\sup_{1 \le t \le (k+1)} \left| \int_1^t d\tilde{B}(t') \right| > k\gamma \right) \le 4e^{-k\gamma^2/2}. \tag{2.157}$$

We also have

$$\begin{aligned} &P\left(\left| \int_0^1 d\tilde{B}(t') \right| < \gamma \text{ and } \sup_{1 \le t \le (k+1)} \left| \int_1^t d\tilde{B})t') \right| < k\gamma \quad \text{for } k = 1, 2, ... \right) \\ &\ge 1 - P\left(\left| \int_0^1 d\tilde{B}(t') \right| \ge \gamma \right) \sum_{k=1}^{\infty} P\left(\sup_{1 \le t \le (k+1)} \left| \int_1^t d\tilde{B})t') \right| > k\gamma \right) \\ &\ge 1 - 4\frac{e^{-\gamma^2/2}}{1 - e^{-\gamma^2/2}} \ge \frac{1 - e^{-\gamma^2/4}}{1 - e^{-\gamma^2/2}} = [1 + e^{-\gamma^2/4}]^{-1}, \end{aligned} \tag{2.158}$$

when $1 - 5e^{-\gamma^2/2} > 1 - e^{-\gamma^2/4}$ which holds if $\gamma > 2\sqrt{\ln(5)}$.

Now we try to bound the first probability on the right hand side of (2.154). To do this, we compare $\tilde{Z}_\varepsilon(t)$ to a Brownian motion with constant drift (See (2.91)). According to (2.57) and the fact that $m_1(s,T)m_2(s,T) \ge \sigma^2(s,T)$[23], we have

$$y(s) \ge \frac{1}{\sigma^2(T)} \left\{ x m_1(s,T)\sigma^2(0,s) + y m_1(0,s)\sigma^2(s,T) + [\sigma^2(0,s) - m_2(0,s)]\sigma^2(s,T) \right\}, \tag{2.159}$$

Since the function $s \to \sigma^2(0,s) - m_2(0,s)$ is an increasing function, we conclude that there exists a constant $C_1 > 0$ depending only on AT such that for $0 \le t \le a\varepsilon$,

$$\tilde{y}(t) \ge \frac{\lambda\varepsilon + \mu_\varepsilon t}{1 + C_1 a\varepsilon/T} \text{ where } \mu_\varepsilon = \frac{m_1(T)y}{\sigma^2(T)} + \frac{\sigma^2(0, s(a\varepsilon)) - m_2(0, s(a\varepsilon))}{\sigma^2(T)}. \tag{2.160}$$

Therefore, following (2.153), for $0 \le t \le a\varepsilon$, there is a constant $C_2 > 0$ depending

[23]This can be shown based on the assumption that $A(s) \ge 0$, $0 \le s \le T$.

only on AT such that

$$\tilde{Z}_\varepsilon(t) \geq Z_\varepsilon(t) \text{ with } Z_\varepsilon(0) = \frac{\lambda\varepsilon}{1 + C_2 a\varepsilon/T}, \mu = \frac{\mu_\varepsilon}{1 + C_2 a\varepsilon/T} - C_2 c_1 \frac{y}{T}. \tag{2.161}$$

Therefore the first probability of (2.154) is bounded below by

$$P\left(Z_\varepsilon(t) > 0, 0 < t \leq a\varepsilon; |Z_\varepsilon(a\varepsilon) - Z_\varepsilon(0) - a\varepsilon\mu| < \gamma\varepsilon\sqrt{a}\middle| Z_\varepsilon(0) = \frac{\lambda\varepsilon}{1 + C_2 a\varepsilon/T}\right). \tag{2.162}$$

To bound the probability in (2.162), we assume that the constant $c_1 > 0$ in (2.151) is small enough such that $\mu > 0$ and $\gamma < \mu\sqrt{a}$. Then using the Dirichlet Green's function as in (2.47), we have

$$\begin{aligned}
&P\left(Z_\varepsilon(t) > 0, 0 < t \leq a\varepsilon; |Z_\varepsilon(a\varepsilon) - Z_\varepsilon(0) - a\varepsilon\mu| < \gamma\varepsilon\sqrt{a}\middle| Z_\varepsilon(0) = \lambda'\varepsilon\right) \\
&= \int_{\lambda'\varepsilon + a\varepsilon\mu - \gamma\varepsilon\sqrt{a}}^{\lambda'\varepsilon + a\varepsilon\mu + \gamma\varepsilon\sqrt{a}} G_{\varepsilon,D}(x, \lambda'\varepsilon, 0, a\varepsilon)dx \\
&= \left\{1 - e^{-2\mu\lambda'}\right\} \frac{1}{\sqrt{2\pi}} \int_{2\lambda'/\sqrt{a}-\gamma}^{\gamma} e^{-z^2/2}dz + \frac{1}{\sqrt{2\pi}} \int_{-\gamma}^{2\lambda'/\sqrt{a}-\gamma} e^{-z^2/2}dz \\
&\quad - e^{-2\mu\lambda'} \frac{1}{\sqrt{2\pi}} \int_{\gamma}^{2\lambda'/\sqrt{a}+\gamma} e^{-z^2/2}dz \\
&\geq \left\{1 - e^{-2\mu\lambda'}\right\} \frac{1}{\sqrt{2\pi}} \int_{-\gamma}^{\gamma} e^{-z^2/2}dz.
\end{aligned} \tag{2.163}$$

The last inequality is true since

$$\left(\int_{-\gamma}^{2\lambda'/\sqrt{a}-\gamma} - \int_{\gamma}^{2\lambda'/\sqrt{a}+\gamma}\right) e^{-z^2/2}dz \geq 0.$$

We take $\lambda' = \lambda/[1 + C_2 a\varepsilon/T]$ in (2.163) and choose $\alpha = 1/2$, whence the drift μ

satisfies the inequality

$$\mu \geq \frac{1}{1+C_3\sqrt{\varepsilon/T}}\left\{1-\frac{m_2(T)}{\sigma^2(T)}+\frac{m_1(T)y}{\sigma^2(T)}\right\}-C_3c_1\frac{y}{T}-C_3\left(\frac{\varepsilon}{T}\right)^{1/2}\,[24],$$

for some constant $C_3>0$ depending only on AT. This is consistent with the condition $\gamma<\mu\sqrt{a}$ [25]. We note that since ε is small, without loss of generality, we assume $\varepsilon/T\leq 1$, then there exists a constant C_4 depending only on AT such that

$$\frac{1}{1+C_3\sqrt{\varepsilon/T}}\geq\frac{1}{1+C_4(\varepsilon/T)^{1/8}}.$$

Now we choose $c_1=c(\varepsilon/T)^{1/8}$ [26] where $c>0$ depends only on AT. Since $\gamma\geq 5$, $y>5T\left(\frac{\varepsilon}{T}\right)^{\alpha/2}$, which implies

$$\left(1-\frac{1}{1+C_4(\varepsilon/T)^{1/8}}\right)\cdot\frac{m_1(T)y}{\sigma^2(T)}>c\left(\frac{\varepsilon}{T}\right)^{\alpha/2+1/8}=c\left(\frac{\varepsilon}{T}\right)^{3/8}>c\left(\frac{\varepsilon}{T}\right)^{1/2}. \quad (2.165)$$

Thus $\left(\frac{\varepsilon}{T}\right)^{1/2}$ can be absorbed by $\left(1-\frac{1}{1+C_4(\varepsilon/T)^{1/8}}\right)\frac{m_1(T)y}{\sigma^2(T)}$. Also $c_1\frac{y}{T}=c(\varepsilon/T)^{1/8}\frac{y}{T}$ can be absorbed by $(1-\frac{1}{1+C_4(\varepsilon/T)^{1/8}})\frac{m_1(T)y}{\sigma^2(T)}$. Therefore, by choosing c small enough, we have

$$\mu\geq\frac{1}{1+C_4(\varepsilon/T)^{1/8}}\left\{1-\frac{m_2(T)}{\sigma^2(T)}+\frac{m_1(T)y}{\sigma^2(T)}\right\}, \quad (2.166)$$

where C_4 depends only on AT. If $\gamma>2\sqrt{\ln(5)}$, the inequality (2.150) follows from (2.158), (2.163), (2.166) and the fact that $\frac{1}{\sqrt{2\pi}}\int_{-\gamma}^{\gamma}e^{-z^2/2}dz>[1+e^{-\gamma^2/4}]^{-1}$ for $\gamma>5$. $\square$

[24] We get $C_3(\varepsilon/T)^{1/2}$ from

$$\sigma^2(0,s(a\varepsilon))-m_2(0,s(a\varepsilon))>\frac{\sigma^2(0,T)-m_2(0,T)}{1+Ca\varepsilon/T}>$$
$$[\sigma^2(0,T)-m_2(0,T)]\cdot\left(1-C'\left(\frac{\varepsilon}{T}\right)^{1/2}\right)>[\sigma^2(0,T)-m_2(0,T)]-C''\left(\frac{\varepsilon}{T}\right)^{1/2}.$$

[25] Therefore $Z_\varepsilon(a\varepsilon)$ in (2.162) is greater than or equal to 0.

[26] We will see below from (2.165) that we only need to choose $c_1=c(\varepsilon/T)^\theta$, where $\theta<1/4$.

2.4 Classical limit theorem

In this section, we will show that the functions $w_\varepsilon(x,t)$ and $\Lambda_\varepsilon(t)$ for the diffusive case will converge to the corresponding functions for the classical case as $\varepsilon \to 0$, where

$$w_\varepsilon(x,t) = \int_x^\infty c_\varepsilon(x',t)dx', \quad h_\varepsilon(x,t) = \int_x^\infty w_\varepsilon(x',t)dx'. \tag{2.167}$$

We note that in order to show $d\Lambda_\varepsilon(t)/dt \to d\Lambda_0(t)/dt$, we need to justify that $\lim_{\varepsilon\to 0} \frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,t)}{\partial x} = c_0(0,t)$ for any $t > 0$.

Lemma II.23. *Let $c_\varepsilon(x,t), \Lambda_\varepsilon(t)$, $0 < x, t < \infty$, be the solution to (2.22), (2.23) with non-negative initial data $c_0(x)$, $0 < x < \infty$, which is a locally integrable function*[27] *satisfying*

$$\int_0^\infty (1+x)c_0(x)dx < \infty, \quad \int_0^\infty xc_0(x)dx = 1. \tag{2.168}$$

Then,

(i) There are positive constants C_1, C_2 depending only on T and $c_0(\cdot)$ such that

$$C_1 \le \Lambda_\varepsilon(t) \le C_2 \quad \text{for } 0 < \varepsilon \le 1, 0 \le t \le T. \tag{2.169}$$

(ii) The set of functions $\{\Lambda_\varepsilon : [0,T] \to \mathbf{R} : 0 < \varepsilon \le 1\}$ form an equicontinuous family.

(iii) Denote by $c_0(x,t)$, $\Lambda_0(t)$, $0 < x, t < \infty$, the solution to the Carr-Penrose equations (2.22), (2.23) with $\varepsilon = 0$ and initial data $c_0(x)$, $0 < x < \infty$. Then for all $x, t \ge 0$

$$\lim_{\varepsilon\to 0} w_\varepsilon(x,t) = w_0(x,t) \tag{2.170}$$

$$\lim_{\varepsilon\to 0} \Lambda_\varepsilon(t) = \Lambda_0(t). \tag{2.171}$$

[27] Integrable on any compact subset of $[0,\infty)$.

These limits are uniform for (x,t) in any finite rectangle $0 < x \leq x_0, 0 < t \leq T$.

Proof. *(i)* It follows (2.26) that $\Lambda_\varepsilon(t)$ is an increasing function, thus $\Lambda_\varepsilon(t)$ has a lower bound $\Lambda_\varepsilon(t) \geq \Lambda_\varepsilon(0) = 1/w(0,0)$.

For the upper bound, we first prove it in *the classical case* $\varepsilon = 0$. The solution to the classical CP equation (2.22) with $\varepsilon = 0$ is given by $w_0(x,t) = w_0(F(x,t),0)$, where $F(x,t)$ is the linear function as in (2.12). Since $\int_0^\infty xc_0(x)dx / \int_0^\infty c_0(x)dx = \Lambda_0(0)$, it follows that

$$\int_{\Lambda_0(0)/2}^\infty xc_0(x)dx \geq \frac{\Lambda_0(0)}{2}\int_0^\infty c_0(x)dx = \frac{1}{2}\int_0^\infty xc_0(x)dx = \frac{1}{2}. \tag{2.172}$$

Hence there is a positive constant $1/C_2$ depending only on $c_0(\cdot)$ such that $w_0(\Lambda_0(0)/2,0) \geq 1/C_2$. Since $F(0,t) < t$, if follows then that $w_0(0,t) = w_0(F(0,t),0) > w_0(t,0) \geq w_0(\Lambda_0(0)/2,0) \geq 1/C_2$ for $0 \leq t \leq \Lambda_0(0)/2$, thus $\Lambda_0(t) \leq C_2$ for $0 \leq t \leq \Lambda_0(0)/2$. Furthermore, $\Lambda_0(t)$ is a continuous function in the interval $0 \leq t \leq \Lambda_0(0)/2$. Since $\Lambda_0(t)$ is an increasing function, similarly we can show that in the interval $t^* \leq t \leq t^* + \Lambda_0(t^*)/2$, where $t^* = \Lambda_0(0)/2$, $\Lambda_0(t) \leq C_3$ and C_3 satisfies $w_0(\Lambda_0(t^*)/2,0) \geq 1/C_3$. This extension covers the interval $0 \leq t \leq T$ in a finite number of steps.

Next, we show the existence of the upper bound for *the diffusive case* when $0 < \varepsilon \leq 1$. To do this, we use the representation

$$w_\varepsilon(x,t) = \int_0^\infty P\left(Y_\varepsilon(t) > x; \inf_{0\leq s\leq t} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(0) = y\right) c_0(y)dy, \tag{2.173}$$

where $Y_\varepsilon(s)$ is the solution to the SDE (2.31) with $b(y,s) = y/\Lambda_\varepsilon(s) - 1$. We know the solution for $Y_\varepsilon(\cdot)$ is given by (2.38), with $A(s) = 1/\Lambda_\varepsilon(s)$. We define a function

$$\tilde{Y}_\varepsilon(s) = \exp\left[\int_0^s \frac{1}{\Lambda_\varepsilon(s')}ds'\right] y - \int_0^s \exp\left[\int_{s'}^s \frac{1}{\Lambda_\varepsilon(s'')}ds''\right] ds'. \tag{2.174}$$

The difference between $Y_\varepsilon(s)$ and $\tilde{Y}_\varepsilon(s)$ is only a Brownian motion term:

$$Y_\varepsilon(s) - \tilde{Y}_\varepsilon(s) = \sqrt{\varepsilon} \int_0^s \exp\left[\int_{s'}^s \frac{1}{\Lambda(s'')} ds''\right] dB(s').$$

Since $\Lambda_\varepsilon(s) \geq \Lambda_\varepsilon(0) = \Lambda_0(0)$, it follows that for any $\delta > 0, t > 0$ there is a positive constant p_1 depending only on $\delta, t, \Lambda_0(0)$ such that

$$P\left(\inf_{0\leq s\leq t}[Y_\varepsilon(s) - \tilde{Y}_\varepsilon(s)] \geq -\delta\right) \geq p_1 \quad \text{for } 0 < \varepsilon \leq 1. \tag{2.175}$$

Because $Y_\varepsilon(s)$ and $\tilde{Y}_\varepsilon(s)$ are close with a large probability, we can choose δ appropriately that there is a positive constant p_2 depending only on $\Lambda_0(0)$ such that if $0 < \varepsilon \leq 1$ then

$$P\left(Y_\varepsilon(t) > 0; \inf_{0\leq s\leq t} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(0) = y\right) \geq p_2 \quad \text{for } t = \Lambda_0(0)/2, y \geq \Lambda_0(0)/2. \tag{2.176}$$

Actually, the choice of δ can be $\min_{0\leq s\leq \Lambda_0(0)/2} \tilde{Y}_\varepsilon(s)$, then p_2 is at least p_1.[28] Therefore, following (2.173), we have

$$w_\varepsilon(0,t) \geq \int_{\Lambda_0(0)/2}^\infty p_2 c_0(y) dy = p_2 w_0(\Lambda_0(0)/2, 0) \geq 1/C_4 \quad \text{for } t = \Lambda_0(0)/2, \tag{2.178}$$

where C_4 only depends on the initial data $c_0(\cdot)$. The same method as for the classical case can be adopted to extend the interval for the upper bound to all T as in the previous paragraph.

(ii) Due to the relationship between Λ_ε and w_ε, we start with the continuity of

[28]We should be aware that under the condition $y \geq \Lambda_0(0)/2$, it is always true that $\tilde{Y}_\varepsilon(s) > 0$ for $0 \leq s \leq \Lambda_0(0)/2$. Actually,

$$\tilde{Y}_\varepsilon(s) = \exp\left[\int_0^s \frac{1}{\Lambda_\varepsilon(s')ds'}\right] \left\{y - \int_0^s \exp\left[-\int_0^{s'} \frac{1}{\Lambda_\varepsilon(s'')} ds''\right] ds'\right\} = \exp\left[\int_0^s \frac{1}{\Lambda_\varepsilon(s')ds'}\right] \{y - s\}. \tag{2.177}$$

$w_\varepsilon(x,t)$.

Following (2.173), we have

$$w_\varepsilon(x,t) - w_\varepsilon(x+\Delta x, t) \le P(x < X + \sqrt{\varepsilon} Z < x + \Delta x) \int_0^\infty c_0(y)dy, \qquad (2.179)$$

where

$$X = \exp\left[\int_0^t \frac{1}{\Lambda_\varepsilon(s)} ds\right] Y - \int_0^t \exp\left[\int_{s'}^t \frac{1}{\Lambda_\varepsilon(s'')} ds''\right] ds', Z = \int_0^t \exp\left[\int_s^t \frac{1}{\Lambda(s')} ds'\right] dB(s), \qquad (2.180)$$

and Y is a random variable with probability density function $c_0(y)/\int_0^\infty c_0(y)dy$. We notice that Z is a Gaussian distribution with variance $\sigma^2 \le \int_0^t \exp\left(2\int_s^t \frac{1}{\Lambda_0(0)} ds'\right) ds$. We also have

$$\begin{aligned} P(x < X + \sqrt{\varepsilon} Z < x + \Delta x) &= P(x - \kappa\Delta x < X < x + (\kappa+1)\Delta x; |\sqrt{\varepsilon} Z| < \kappa \Delta x) \\ &\quad + P(x < X + \sqrt{\varepsilon} Z < x + \Delta x; |\sqrt{\varepsilon} Z| \ge \kappa\Delta x). \qquad (2.181) \end{aligned}$$

To estimate the first term,

$$\begin{aligned} &P(x - \kappa\Delta x < X < x + (\kappa+1)\Delta x; |\sqrt{\varepsilon} Z| < \kappa\Delta x) \\ &\qquad \le P(x - \kappa\Delta x < X < x + (\kappa+1)\Delta x) \qquad (2.182) \end{aligned}$$

where C is a constant depending on $c_0(x)$. The expression in (2.182) converges to 0 as $(2\kappa+1)\Delta x$ approaches to 0 due to the relation between X and Y in (2.180), the probability density function of Y and condition (2.168). For the second term,

$$
\begin{aligned}
& P(x < X + \sqrt{\varepsilon} Z < x + \Delta x; |\sqrt{\varepsilon} Z| \geq \kappa \Delta x) \\
= & \int_0^\infty P(X \in dy) P(|\sqrt{\varepsilon} Z| \geq \kappa \Delta x; x - y < \sqrt{\varepsilon} Z < x - y + \Delta x) \\
\leq & \int_0^\infty P(X \in dy) \cdot \sup_{a \geq (\kappa - 1)\Delta x} P(a \leq |\sqrt{\varepsilon} Z| \leq a + \Delta x) \\
< & \ C' \Delta x \frac{1}{\sqrt{\varepsilon \sigma^2}} \exp\left(-\frac{\kappa^2 \Delta x^2}{2 \varepsilon \sigma^2} \right) \\
< & \ \frac{C'}{\kappa} \max_{z > 0} \left(z \cdot e^{-z^2/2} \right),
\end{aligned}
\tag{2.183}
$$

where C' is a positive constant. Therefore, by choosing $\kappa = 1/\sqrt{\Delta x}$, both $(2\kappa + 1)\Delta x$ and (2.183) converges to 0 as Δx converges to 0. Thus we have shown that w_ε is uniformly continuous in terms of x.

Next, we observe that for $\Delta t > 0$ there exists $x(\Delta t)$ independent of ε in the interval $0 < \varepsilon \leq 1$ such that $\lim_{\Delta t \to 0} x(\Delta t) = 0$ and

$$
\begin{aligned}
& P\left(Y_\varepsilon(t + \Delta t) > 0; \inf_{0 \leq s \leq t + \Delta t} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(0) = y \right) \geq \\
& \quad [1 - \Delta t] P\left(Y_\varepsilon(t) > x(\Delta t); \inf_{0 \leq s \leq t} Y_\varepsilon(s) > 0 \middle| Y_\varepsilon(0) = y \right) \quad \text{for } y \geq 0,\ 0 < \varepsilon \leq 1.
\end{aligned}
\tag{2.184}
$$

It follows from (2.173), (2.184) that $w_\varepsilon(0, t+\Delta t) \geq [1-\Delta t] w_\varepsilon(x(\Delta t), t)$ for $0 < \varepsilon \leq 1$. Using the continuity of $w_\varepsilon(x,t)$ in terms of x, we conclude that $\lim_{\Delta t \to 0} w_\varepsilon(0, t+\Delta t) \geq w_\varepsilon(0,t)$ and the limit is uniform for $0 < \varepsilon \leq 1$. On the other side, $w_\varepsilon(0, t + \Delta t) = 1/\Lambda_\varepsilon(t + \Delta t) \leq 1/\Lambda_\varepsilon(t) = w_\varepsilon(0,t)$, which leads to the result that $\lim_{\Delta t \to 0} w_\varepsilon(0, t + \Delta t) = w_\varepsilon(0,t)$. Therefore, the function $\Lambda_\varepsilon(t)$ is continuous, and in fact the family of functions $\Lambda_\varepsilon(t)$, $0 < \varepsilon \leq 1$, is equicontinuous.

(iii) Due to Ascoli-Arzela theorem and the fact that the family of functions $w_\varepsilon(x,t)$, $\Lambda_\varepsilon(\cdot)$, $0 < \varepsilon \leq 1$, are equicontinuous, there exists a subsequence $\{\varepsilon_n\}$ with

$\varepsilon_n \to 0$, as $n \to \infty$, such that $w_{\varepsilon_n}(x,t)$ and $\Lambda_{\varepsilon_n}(\cdot)$ converge uniformly respectively. The limit satisfy the condition $w_0(x,t) = w_0(F(x,t),0)$ and the conservation law (2.23) continues to hold for $\varepsilon = 0$. Thus the limits are the solution to the classical model. Since the solution to the classical model is unique, it follows that for all $\varepsilon \to 0$, (2.170) and (2.171) hold true. The uniformity of the limits follows by similar argument. $\square$

Lemma II.24. *Let $c_\varepsilon(x,t), \Lambda_\varepsilon(t)$, $0 < x,t < \infty$, and $c_0(x)$, $0 < x < \infty$, be as in Lemma (II.23) and (2.168). If $c_0(\cdot)$ is a continuous function then*

$$\lim_{\varepsilon \to 0} \frac{\varepsilon}{2} \frac{\partial c_\varepsilon(0,T)}{\partial x} = c_0(0,T) \quad \textit{for any } T > 0. \tag{2.185}$$

Proof. We use the identity

$$\frac{\varepsilon}{2} \frac{\partial c_\varepsilon(0,T)}{\partial x} = \lim_{\lambda \to 0} \frac{c_\varepsilon(\lambda\varepsilon, T)}{2\lambda} \tag{2.186}$$

and the representation for $c_\varepsilon(\lambda\varepsilon, T)$

$$c_\varepsilon(\lambda\varepsilon, T) = \int_0^\infty G_{\varepsilon,D}(\lambda\varepsilon, y, 0, T) c_0(y) dy, \tag{2.187}$$

where $G_{\varepsilon,D}$ is the Dirichlet Green's function corresponding to the drift $b(y,t) = y/\Lambda_\varepsilon(t) - 1$. Let $\sigma_\varepsilon^2(T), m_{1,\varepsilon}(T), m_{2,\varepsilon}(T)$ be the functions in (2.39) and (2.40) with $A(s) = 1/\Lambda_\varepsilon(s), 0 \le s \le T$. Therefore we have

$$\begin{aligned} &\frac{\varepsilon}{2} \frac{\partial c_\varepsilon(0,T)}{\partial x} \\ &\quad \le \int_0^\infty \left\{ 1 - \frac{m_{2,\varepsilon}(T)}{\sigma_\varepsilon^2(T)} + \frac{m_{1,\varepsilon}(T) y}{\sigma_\varepsilon^2(T)} + C\Gamma\left(\frac{\varepsilon}{T}, \frac{y}{T}\right) \left[1 + \frac{y}{T}\right] \right\} G_\varepsilon(0,y,0,T) c_0(y) dy, \end{aligned}$$

where the constant C depends only on $T/\Lambda_0(0)$. When $\varepsilon \to 0$, $G_\varepsilon(0,y,0,T) \to$

$\frac{1}{m_{1,0}(T)}\delta\left(y, \frac{m_{2,0}(T)}{m_{1,0}(T)}\right)$. While when $y = \frac{m_{2,0}(T)}{m_{1,0}(T)}$,

$$\lim_{\varepsilon\to 0}\left\{1 - \frac{m_{2,\varepsilon}(T)}{\sigma_\varepsilon^2(T)} + \frac{m_{1,\varepsilon}(T)y}{\sigma_\varepsilon^2(T)} + C\Gamma\left(\frac{\varepsilon}{T}, \frac{y}{T}\left[1+\frac{y}{T}\right]\right)\right\} = 1.$$

Therefore, by *reverse Fatou's lemma* (See [15, page 97]),

$$\limsup_{\varepsilon\to 0}\frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,T)}{\partial x} \le \frac{1}{m_{1,0}(T)} c_0\left(\frac{m_{2,0}(T)}{m_{1,0}(T)}\right), \tag{2.188}$$

provided the function $c_0(y)$, $y > 0$, is continuous at $y = m_{2,0}(T)/m_{1,0}(T)$.

On the other side, we can obtain a lower bound of $\frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,T)}{\partial x}$ by using Lemma II.22. Thus we have

$$\begin{aligned}\left[1 + C(\varepsilon/T)^{1/8}\right]\frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,T)}{\partial x} \ge \\ \int_{5(\varepsilon/T)^{1/8}T/c}^{\infty}\left[1+\exp\left(-c^2(T/\varepsilon)^{1/4}(y^2/4T^2)\right)\right]^{-2}\left\{1 - \frac{m_{2,\varepsilon}(T)}{\sigma_\varepsilon^2(T)} + \frac{m_{1,\varepsilon}(T)y}{\sigma_\varepsilon^2(T)}\right\} \\ \cdot G_\varepsilon(0,y,0,T)c_0(y)dy,\end{aligned} \tag{2.189}$$

where the constants $C, c > 0$ depend only on $T/\Lambda_0(0)$. We conclude that

$$\liminf_{\varepsilon\to 0}\frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,T)}{\partial x} \ge \frac{1}{m_{1,0}(T)} c_0\left(\frac{m_{2,0}(T)}{m_{1,0}(T)}\right), \tag{2.190}$$

provided the function $c_0(y)$, $y > 0$ is continuous at $y = m_{2,0}(T)/m_{1,0}(T)$. Finally, since $w_0(x,t) = w_0(F(x,t))$, by differentiating this equation with respect to x at $x = 0$, we obtain that

$$c_0(0,T) = \frac{1}{m_{1,0}(T)} c_0\left(\frac{m_{2,0}(T)}{m_{1,0}(T)}\right).$$

Thus (2.185) has been proved. □

In light of Lemma II.3, II.24, (2.26) and the fact that $\lim_{\varepsilon\to 0}\Lambda_\varepsilon(t) = \Lambda_0(t)$, we conclude that

$$\lim_{\varepsilon\to 0}\frac{d\Lambda_\varepsilon(t)}{dt}=\frac{d\Lambda_0(t)}{dt}. \tag{2.191}$$

2.5 Upper bound on the coarsening rate

In this section, we show there is an upper bound on the coarsening rate if the initial data satisfies a convexity condition. We first recall the definition of $\beta_\varepsilon(x,t)$ as in (2.17),

$$\beta_\varepsilon(x,t)=\frac{c_\varepsilon(x,t)h_\varepsilon(x,t)}{w_\varepsilon(x,t)^2}, \tag{2.192}$$

where $w_\varepsilon, h_\varepsilon$ are given by (2.167). We observe that $h_\varepsilon(x,t)$ is log concave in $x>0$ if and only if $\sup_{x>0}\beta_\varepsilon(x,t)\le 1$. Actually, if we assume that $h_\varepsilon(x,t)=\exp[-q_\varepsilon(x,t)]$, then

$$\beta_\varepsilon(x,t)=1-\frac{\partial^2 q_\varepsilon/\partial x^2}{(\partial q_\varepsilon/\partial x)^2}. \tag{2.193}$$

It is clear that $\sup_{x>0}\beta_\varepsilon(x,t)\le 1$ if and only if $\partial^2 q_\varepsilon/\partial x^2\ge 0$, i.e., $h_\varepsilon(x,t)$ is log concave in $x>0$.

We shall show that if $w_\varepsilon(x,0)$ and $h_\varepsilon(x,0)$ are log concave respectively, then so are $w_\varepsilon(x,t)$ and $h_\varepsilon(x,t)$ for all $t>0$. Due to the PDE (2.22) that $c_\varepsilon(x,t)$ satisfies, we can derive corresponding PDEs for w_ε and h_ε,

$$\frac{\partial w_\varepsilon}{\partial t}+\frac{\partial}{\partial x}\left[\left(\frac{x}{\Lambda_\varepsilon(t)}-1\right)w_\varepsilon\right]=\frac{w_\varepsilon}{\Lambda_\varepsilon(t)}+\frac{\varepsilon}{2}\frac{\partial^2 w_\varepsilon}{\partial x^2}. \tag{2.194}$$

$$\frac{\partial h_\varepsilon}{\partial t}+\frac{\partial}{\partial x}\left[\left(\frac{x}{\Lambda_\varepsilon(t)}-1\right)h_\varepsilon\right]=\frac{2h_\varepsilon}{\Lambda_\varepsilon(t)}+\frac{\varepsilon}{2}\frac{\partial^2 h_\varepsilon}{\partial x^2}. \tag{2.195}$$

If there is no Dirichlet boundary condition, then

$$w_\varepsilon(x,T)=\exp\left[\int_0^T\frac{ds}{\Lambda_\varepsilon(s)}\right]\int_{-\infty}^{\infty}G_\varepsilon(x,y,0,T)w_\varepsilon(y,0)dy,$$
$$h_\varepsilon(x,T)=\exp\left[2\int_0^T\frac{ds}{\Lambda_\varepsilon(s)}\right]\int_{-\infty}^{\infty}G_\varepsilon(x,y,0,T)h_\varepsilon(y,0)dy.$$

If $w_\varepsilon(x,0), h_\varepsilon(x,0)$ are log concave functions then the representation of $w_\varepsilon(x,T), h_\varepsilon(x,T)$ are convolutions of two log concave functions. It follows from the Prékopa-Leindler theorem that $w_\varepsilon(x,T), h_\varepsilon(x,T)$ are also log concave [33]. To show the corresponding result for the Dirichlet problem, we use the method of Korevaar [27].

Lemma II.25. *Suppose $c_0 : [0,\infty) \to \mathbf{R}^+$ satisfies (2.168) and $c_\varepsilon(x,t)$, $x \geq 0, t > 0$ is the solution to (2.22), (2.23). If the function $w_\varepsilon(x,t)$ is log concave in x at $t = 0$, then it is log concave in x for all $t > 0$.*

Proof. The function $c_\varepsilon(x,t)$ is C^∞ in the domain $\{x, t > 0\}$ and continuous in the domain $\{x \geq 0, t > 0\}$ with $c_\varepsilon(0,t) = 0$ for $t > 0$. We define the function v_ε by

$$v_\varepsilon(x,t) = -\frac{\partial}{\partial x} \log w_\varepsilon(x,t) \quad \text{for } x, t > 0.$$

In order to show that $w_\varepsilon(x,t)$ is log concave, we only need to show that $v_\varepsilon(x,t)$ is increasing in x. By plugging the expression $w_\varepsilon(x,t) = \exp[-\int v_\varepsilon(x,t)dx]$ into (2.194), we obtain a PDE for $v_\varepsilon(x,t)$

$$\frac{\partial v_\varepsilon(x,t)}{\partial t} + \left[\frac{x}{\Lambda_\varepsilon(t)} - 1 + \varepsilon v_\varepsilon(x,t)\right] \frac{\partial v_\varepsilon(x,t)}{\partial x} + \frac{v_\varepsilon(x,t)}{\Lambda_\varepsilon(t)} = \frac{\varepsilon}{2} \frac{\partial^2 v_\varepsilon(x,t)}{\partial x^2}. \tag{2.196}$$

Since $c_\varepsilon(0,t) = 0, t > 0$, it follows the definition of v_ε that it satisfies the Dirichlet condition $v_\varepsilon(0,t) = 0$ for $t > 0$. We consider a diffusion process $X_\varepsilon(\cdot)$ run backwards in time, which is the solution to the stochastic equation (2.36) with $b(x,s)$, $x, s > 0$, given by the formula

$$b(x,s) = \frac{x}{\Lambda_\varepsilon(s)} - 1 + \varepsilon v_\varepsilon(x,s).$$

For x let $\tau_{\varepsilon,x} < T$ be the first hitting time at 0 of $X_\varepsilon(s), s < T$, with $X_\varepsilon(T) = x$.

Then as in (2.37), we have

$$v_\varepsilon(x,T) = \exp\left\{-\int_0^T \frac{dt}{\Lambda_\varepsilon(t)}\right\} E\left[\frac{c_0(X_\varepsilon(0))}{w_0(X_\varepsilon(0))}; \tau_{\varepsilon,x} \le 0 \Big| X_\varepsilon(T) = x\right]^{29}. \tag{2.197}$$

Suppose now that $0 < x_1 < x_2$ and $X_{\varepsilon,j}(s), s \le T$, is the solution to (2.36) with $X_{\varepsilon,j}(T) = x_j$, $j = 1,2$. By taking the same copy of white noise for $X_{\varepsilon,1}(\cdot)$ and $X_{\varepsilon,2}(\cdot)$, it is clear that $X_{\varepsilon,1}(s) \le X_{\varepsilon,2}(s)$ for $s \le T$, hence $\tau_{\varepsilon,1} \ge \tau_{\varepsilon,2}$, where $\tau_{\varepsilon,j}$ denotes the corresponding first hitting time at 0 for $X_{\varepsilon,j}$, $j = 1,2$. Due to the log concavity of $w_\varepsilon(\cdot,0)$, $c_0(x)/w_0(x)$ is an increasing function. We conclude from (2.197) that $v_\varepsilon(x_1,T) \le v_\varepsilon(x_2,T)$. Thus far, we have proved that $v_\varepsilon(x,T)$ is an increasing function of $x > 0$, so $w_\varepsilon(x,T)$ is log concave in $x > 0$. □

Lemma II.26. *Suppose $c_0 : [0,\infty) \to \mathbf{R}^+$ satisfies (2.168) and $c_\varepsilon(x,t)$, $x \ge 0, t > 0$ is the solution to (2.22), (2.23). If the function $h_\varepsilon(x,t)$ is log concave in x at $t = 0$, then it is log concave in x for all $t > 0$.*

Proof. Let $v_\varepsilon(x,t) = -\frac{\partial}{\partial x}\log h_\varepsilon(x,t)$ (which is different from the definition in Lemma II.25), then following (2.195), we have

$$\frac{\partial v_\varepsilon(x,t)}{\partial t} + \left[\frac{x}{\Lambda_\varepsilon(t)} - 1 + \varepsilon v_\varepsilon(x,t)\right]\frac{\partial v_\varepsilon(x,t)}{\partial x} + \frac{v_\varepsilon(x,t)}{\Lambda_\varepsilon(t)} = \frac{\varepsilon}{2}\frac{\partial^2 v_\varepsilon(x,t)}{\partial x^2}. \tag{2.198}$$

By differentiating this equation with respect to x, we can obtain a PDE for the function $u_\varepsilon(x,t) = \frac{\partial}{\partial x}v_\varepsilon(x,t)$,

$$\frac{\partial u_\varepsilon(x,t)}{\partial t} + \left[\frac{x}{\Lambda_\varepsilon(t)} - 1 + \varepsilon v_\varepsilon(x,t)\right]\frac{\partial u_\varepsilon(x,t)}{\partial x} + \frac{2u_\varepsilon(x,t)}{\Lambda_\varepsilon(t)} + \varepsilon u_\varepsilon(x,t)^2 = \frac{\varepsilon}{2}\frac{\partial^2 u_\varepsilon(x,t)}{\partial x^2}. \tag{2.199}$$

At the same time, we observe that

$$u_\varepsilon(x,t) = v_\varepsilon(x,t)^2[1 - c_\varepsilon(x,t)h_\varepsilon(x,t)/w_\varepsilon(x,t)^2]. \tag{2.200}$$

[29] Notice that $v_\varepsilon(x,t) = c_\varepsilon(x,t)/w_\varepsilon(x,t)$

Since $\lim_{x\to 0} c_\varepsilon(x,t) = 0$ for $t > 0$, it follows that $\liminf_{x\to 0} u_\varepsilon(x,t) \geq 0$ for $t > 0$. If $h_\varepsilon(x,0)$ is log concave in $x > 0$ then the initial data $u_\varepsilon(x,0)$, $x > 0$ is non-negative. According to *the maximum principle*[30], $u_\varepsilon(x,t)$ is non-negative for all $x, t > 0$, and hence $h_\varepsilon(x,t)$ is a log concave function of x for all $t > 0$. $\square$

Lemma II.27. *Let $c_\varepsilon(x,t), \Lambda_\varepsilon(t)$, $0 < x, t < \infty$, be the solution to (2.22), (2.23) with $\varepsilon > 0$ and non-negative initial data $c_0(x)$, $0 < x < \infty$, which is locally integrable function satisfying (2.168). Then $\lim_{t\to\infty} \Lambda_\varepsilon(t) = \infty$.*

Proof. We have already shown that $\Lambda_\varepsilon(t)$ is an increasing function of t. Therefore, it is sufficient to show that if there is an upper bound for $\Lambda_\varepsilon(t)$, i.e., $\Lambda_\varepsilon(t) \leq \Lambda_\infty$ for some finite Λ_∞, then there is a contradiction. To see this, we use the identity

$$\begin{aligned}
\frac{d}{dt}\int_0^\infty x c_\varepsilon(x,t)dx &= \int_0^\infty x\left\{\frac{\partial}{\partial x}\left[1 - \frac{x}{\Lambda_\varepsilon(t)}\right] c_\varepsilon(x,t) + \frac{\varepsilon}{2}\frac{\partial^2 c_\varepsilon(x,t)}{\partial x^2}\right\} dx \\
&= \frac{1}{\Lambda_\varepsilon(t)}\int_0^\infty x c_\varepsilon(x,t)dx - \int_0^\infty c_\varepsilon(x,t)dx,
\end{aligned}$$

where the first equation follows (2.22) and the second integration by parts and Dirichlet boundary condition. Because of the upper bound for $\Lambda_\varepsilon(t)$, we have

$$\begin{aligned}
\frac{d}{dt}\int_0^\infty x c_\varepsilon(x,t)dx \geq& \left(\frac{1}{2\Lambda_\infty}\int_0^\infty x c_\varepsilon(x,t)dx - \int_{2\Lambda_\infty}^\infty c_\varepsilon(x,t)dx\right) \\
&+ \frac{1}{2\Lambda_\infty}\int_0^\infty x c_\varepsilon(x,t)dx - \int_0^{2\Lambda_\infty} c_\varepsilon(x,t)dx \\
\geq& \frac{1}{2\Lambda_\infty}\int_0^\infty x c_\varepsilon(x,t)dx - \int_0^{2\Lambda_\infty} c_\varepsilon(x,t)dx \\
=& \frac{1}{2\Lambda_\infty} - \int_0^{2\Lambda_\infty} c_\varepsilon(x,t)dx \\
\geq& \frac{1}{2\Lambda_\infty} - \int_0^{2\Lambda_\infty} dx \int_0^\infty G_\varepsilon(x,y,0,t)c_0(y)dy,
\end{aligned}$$

[30] See Theorem 1 of [13, page 344]. Though (2.199) is not technically a linear function of u_ε due to the term $\varepsilon u_\varepsilon^2(x,t)$. But we can see it as one with $\varepsilon u_\varepsilon(x,t)$ being a term in front of $u_\varepsilon(x,t)$. If we see (2.199) from the Feynman-Kac Theorem point of view, the term $\varepsilon u_\varepsilon(x,t)$ only influences the discount factor, not the sign of the function.

where the equality follows the conservation law (2.23), and the last inequality follows from the fact that $G_{\varepsilon,D}(x,y,0,t) \le G_\varepsilon(x,y,0,t)$, and G_ε is the function (2.41) with $A(s) = 1/\Lambda_\varepsilon(s)$, $s \ge 0$. Thus

$$\int_0^\infty G_\varepsilon(x,y,0,t)c_0(y)dy \ge \frac{1}{\sqrt{2\pi\varepsilon\sigma^2(t)}}\int_0^\infty c_0(y)dy \ge \frac{1}{\sqrt{2\pi\varepsilon\sigma^2(t)}\Lambda_\varepsilon(0)},$$

and it follows that

$$\frac{d}{dt}\int_0^\infty xc_\varepsilon(x,t)dx \ge \frac{1}{2\Lambda_\infty} - \frac{2\Lambda_\infty}{\sqrt{2\pi\varepsilon\sigma^2(t)}\Lambda_\varepsilon(0)}.$$

Since $\sigma^2(t) \ge t$, we conclude that

$$\lim_{t\to\infty}\int_0^\infty xc_\varepsilon(x,t)dx = \infty,$$

contradicting the conservation law (2.23). Therefore, there is no upper bound for $\Lambda_\varepsilon(t)$ for all t, i.e., $\lim_{t\to\infty}\Lambda_\varepsilon(t) = \infty$. □

Lemma II.28. *Suppose* $c_0 : [0,\infty) \to \mathbf{R}^+$ *satisfies* (2.168) *and* $c_\varepsilon(x,t)$, $x \ge 0, t > 0$ *is the solution to* (2.22), (2.23). *Assume that* $\Lambda_\varepsilon(0) = 1$ *and that the function* $h_\varepsilon(x,0)$ *is log concave in* x. *Then there exist positive universal constants* C, ε_0 *with* $0 < \varepsilon_0 \le 1$ *such that*

$$c_\varepsilon(\lambda\varepsilon,1) \le C\lambda c_\varepsilon(\varepsilon,1) \quad \textit{for } 0 < \varepsilon \le \varepsilon_0, 0 < \lambda \le 1. \tag{2.201}$$

Proof. Let X_0 be the positive random variable with probability density function $c_0(x)/\int_0^\infty c_0(x')dx'$, $x > 0$. Then we see that $\langle X_0\rangle = 1$ due to the assumption $\Lambda_\varepsilon(0) = 1$. We choose a constant $\zeta = 1 + 0.1e^{-1} > 1$. According to Markov inequality,

$$P(X_0 > \zeta) \le 1/\zeta < 1.$$

Since $h_\varepsilon(x,0)$ is log concave in x, $\beta(x,0) \le 1$, and it follows from (29) of [8] that there exists a universal constant δ with $0 < \delta < 1 - 2.2e^{-1}$ such that

$$P(X_0 < \delta) + P(X_0 > \zeta) \leq c < 1^{31}, \tag{2.202}$$

where c is a universal constant. We write

$$c_\varepsilon(\lambda\varepsilon, 1) = \int_0^\infty G_{\varepsilon,D}(\lambda\varepsilon, y, 0, 1)c_0(y)dy = \left[\int_0^\delta + \int_\delta^\zeta + \int_\zeta^\infty\right] G_{\varepsilon,D}(\lambda\varepsilon, y, 0, 1)c_0(y)dy \tag{2.203}$$

We argue that when δ is small the whole integral from 0 to ∞ can be bounded by a universal constant times the integral from δ to ζ. Since the characteristic function

$$F_\varepsilon(x, 1) := \exp\left[-\int_0^1 \frac{1}{\Lambda_\varepsilon(s)}ds\right] x + \int_0^1 \exp\left[-\int_0^s \frac{1}{\Lambda_\varepsilon(s')}ds'\right] ds,$$

and $\Lambda_\varepsilon(\cdot) > 1$, it follows that $1-e^{-1} < F_\varepsilon(0, 1) < 1$. For $0 < \lambda < 1$, we can also choose ε_0 small enough such that $F_\varepsilon(\lambda\varepsilon, 1)$ satisfy the same inequality: $1 - e^{-1} < F_\varepsilon(\lambda\varepsilon, 1) < 1$. Since $G_\varepsilon(\lambda\varepsilon, y, 0, 1)$ has the expression given by (2.41) with the function $A(\cdot)$ in (2.39) replaced by $1/\Lambda_\varepsilon(\cdot)$, the axis of symmetry for $G_\varepsilon(\lambda, y, 0, 1)$ is at $F_\varepsilon(\lambda\varepsilon, 1)$. Because $\delta < 1 - 2.2e^{-1}$,

$$\zeta - F(\lambda\varepsilon, 1) < 1.1e^{-1} < 1.2e^{-1} < F(\lambda\varepsilon, 1) - \delta,$$

thus we always have $\min_{\delta\leq y\leq\zeta} G_\varepsilon(\lambda\varepsilon, y, 0, 1) \geq \max_{0\leq y\leq\delta} G_\varepsilon(\lambda\varepsilon, y, 0, 1)$.

[31] See Remark II.29 for a detailed explanation on the existence of such δ.

According to Lemma II.20, we have

$$\begin{aligned}&\int_0^\delta G_{\varepsilon,D}(\lambda\varepsilon, y, 0, 1)c_0(y)dy\\ &\quad\le \int_0^\delta \lambda\left\{2\left(1-\frac{m_2(1)}{\sigma^2(1)}+\frac{m_1(1)y}{\sigma^2(1)}\right)+C\,(1+y)\right\}G_\varepsilon(\lambda\varepsilon, y, 0, 1)c_0(y)dy\\ &\qquad\le C_1'\lambda\int_0^\delta G_\varepsilon(\lambda\varepsilon, y, 0, 1)c_0(y)dy \le C_1\lambda\int_\delta^\zeta G_\varepsilon(\lambda\varepsilon, y, 0, 1)c_0(y)dy,\end{aligned}$$

where C_1, C_1' are universal constants. Also,

$$\begin{aligned}&\int_\zeta^\infty G_{\varepsilon,D}(\lambda\varepsilon, y, 0, 1)c_0(y)dy\\ &\quad\le \int_\zeta^\infty \lambda\left\{2\left(1-\frac{m_2(1)}{\sigma^2(1)}+\frac{m_1(1)y}{\sigma^2(1)}\right)+C\,(1+y)\right\}G_\varepsilon(\lambda\varepsilon, y, 0, 1)c_0(y)dy.\end{aligned}$$

Since $G_\varepsilon(\lambda\varepsilon, y, 0, 1)$ converges to 0 exponentially fast as $y\to\infty$, the integral above is finite and there exist universal constants C_2, C_2' such that

$$\begin{aligned}\int_\zeta^\infty G_{\varepsilon,D}(\lambda\varepsilon, y, 0, 1)c_0(y)dy &\le C_2'\int_\zeta^\infty (1+y)G_\varepsilon(\lambda\varepsilon, y, 0, 1)c_0(y)dy\\ &\le C_2\int_\delta^\zeta G_\varepsilon(\lambda\varepsilon, y, 0, 1)c_0(y)dy.\end{aligned}$$

The second inequality holds since $\sup_{y\ge\zeta}(1+y)G_\varepsilon(\lambda\varepsilon, y, 0, 1)\le const\cdot\min_{\delta\le y\le\zeta}G_\varepsilon(\lambda\varepsilon, y, 0, 1)$. Further, with $0<\lambda\le 1$, there exists a universal constant C_3 such that $G_\varepsilon(\lambda\varepsilon, y, 0, 1)\le C_3 G_\varepsilon(\varepsilon, y, 0, 1)$ provided $y\le S$. Therefore,

$$c_\varepsilon(\lambda\varepsilon, 1)\le C_4\lambda\int_\delta^\zeta G_\varepsilon(\varepsilon, y, 0, 1)c_0(y)dy \quad \text{for } 0<\varepsilon,\lambda\le 1, \tag{2.204}$$

where $C_4 > 0$ is a universal constant. By applying Lemma II.22,

$$\begin{aligned}
&\int_\delta^\zeta G_\varepsilon(\varepsilon, y, 0, 1)c_0(y)dy \\
&\qquad \le \int_\delta^\zeta [1+e^{-\gamma^2/4}]^2 \left(1-\exp\left[-\frac{2}{1+C\varepsilon^{1/8}}\left\{1-\frac{m_2(1)}{\sigma^2(1)}+\frac{m_1(1)y}{\sigma^2(1)}\right\}\right]\right)^{-1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot G_{\varepsilon,D}(\lambda\varepsilon, y, 0, 1)c_0(y)dy,
\end{aligned}$$

where $\gamma = c(1/\varepsilon)^{1/8}y \ge 5$ as long as we choose a small enough ε_0 (we notice that here the choice of y has a lower bound δ). Since for a fixed δ, there exists a universal constant C_5 such that

$$[1+e^{-\gamma^2/4}]^2 \left(1-\exp\left[-\frac{2}{1+C\varepsilon^{1/8}}\left\{1-\frac{m_2(1)}{\sigma^2(1)}+\frac{m_1(1)y}{\sigma^2(1)}\right\}\right]\right)^{-1} \le C_5, \text{ for } y \ge \delta,$$

thus we have

$$\int_\delta^\zeta G_\varepsilon(\varepsilon, y, 0, 1)c_0(y)dy \le C_5 \int_\delta^\zeta G_{\varepsilon,D}(\varepsilon, y, 0, 1)c_0(y)dy \le C_5 c_\varepsilon(\varepsilon, 1). \tag{2.205}$$

Therefore, (2.201) follows from (2.204) and (2.205). $\square$

Remark II.29. *For a positive random variable X whose probability density function is $c(x)/\int_0^\infty c(x')dx'$, we define $q(x) = -\log(h(x))$, then $h(x) = \exp(-q(x))$, $w(x) = q'(x)h(x)$ and $c(x) = \{[q'(x)]^2 - q''(x)\}\, h(x)$. Therefore*

$$\beta(x) = \frac{c(x)h(x)}{w^2(x)} = 1 - \frac{q''(x)}{[q'(x)]^2}.$$

By solving this differential equation of $q'(x)$, we obtain

$$q'(x) = \left[\frac{1}{q'(0)} - x + \int_0^x \beta(z)dz\right]^{-1}.$$

Thus

$$q(x) = q(0) + \int_0^x dz \Big/ \left[1/q'(0) - z + \int_0^z \beta(z')dz'\right], \quad 0 \le x < \|X\|_\infty{}^{32}.$$

Also because $\langle X\rangle = h(0)/w(0) = 1/q'(0)$, *it follows that*

$$\begin{aligned}\frac{w(x)}{w(0)} = \frac{\langle X\rangle}{\left[\langle X\rangle - x + \int_0^x \beta(z')dz'\right]} \exp\left[-\int_0^x \frac{dz}{\left[\langle X\rangle - z + \int_0^z \beta(z')dz'\right]}\right] \\ = \exp\left[-\int_0^x \frac{\beta(z)dz}{\langle X\rangle - z + \int_0^z \beta(z')dz'}\right]. \quad (2.206)\end{aligned}$$

It is noted that $\|X\|_\infty - \int_0^{\|X\|_\infty} \beta(z')dz' = \langle X\rangle$ *thus* (2.206) *is always positive and smaller than 1 for* $0 \le x \le \|X\|_\infty$. *The second equation of* (2.206) *holds since* $\log[\langle X\rangle - x + \int_0^x \beta(z')dz']$ *satisfies differential equation*

$$\frac{d}{dx}\log[\langle X\rangle - x + \int_0^x \beta(z')dz'] = \frac{\beta(x) - 1}{\langle X\rangle - x + \int_0^x \beta(z')dz'},$$

thus

$$\frac{\langle X\rangle}{\left[\langle X\rangle - x + \int_0^x \beta(z')dz'\right]} = \exp\left[\int_0^x \frac{1 - \beta(z)}{\langle X\rangle - x + \int_0^x \beta(z')dz'}\right].$$

Before (2.202), *we assume* $\langle X_0\rangle = 1$ *and* $\beta(x,0) \le 1$, *thus*

$$w(x)/w(0) \ge 1 \cdot \exp\left[-\int_0^x \frac{dz}{1-z}\right]. \quad (2.207)$$

So there exists a universal $\delta > 0$ *such that when* $x \le \delta$, $w(x)/w(0)$ *large enough such that* (2.202) *holds.*

Lemma II.30. *Suppose the initial data* $c_0(\cdot)$ *for* (2.22), (2.23) *satisfies the conditions of Lemma* (II.28) *and* $0 < \varepsilon \le \varepsilon_0$. *Then there is a universal constant* C *such that* $d\Lambda_\varepsilon(t)/dt \le C$ *for* $t \ge 1$.

[32] $\|X\|_\infty = \sup\{x : c(x) > 0\}$

Proof. Recall that $h_\varepsilon(0,t) = 1$ for all t due to the conservation law, thus according to (2.26), at $t = 1$,

$$\frac{d\Lambda_\varepsilon(1)}{dt} = \frac{\varepsilon}{2}\frac{\partial c_\varepsilon(0,1)}{\partial x} \cdot h_\varepsilon(0,1)\Big/ w_\varepsilon(0,1)^2, \tag{2.208}$$

where

$$\varepsilon\frac{\partial c_\varepsilon(0,1)}{\partial x} = \lim_{\lambda\to 0} \varepsilon\frac{c_\varepsilon(\lambda\varepsilon,1)}{\lambda\varepsilon} \le \lim_{\lambda\to 0}\frac{C_1\lambda\varepsilon c_\varepsilon(\varepsilon,1)}{\lambda\varepsilon} = C_1 c_\varepsilon(\varepsilon,1).$$

Therefore,

$$\frac{d\Lambda_\varepsilon(1)}{dt} \le \frac{C_1 c_\varepsilon(\varepsilon,1)h_\varepsilon(0,1)}{w_\varepsilon(0,1)^2} \le \frac{C_1\beta_\varepsilon(\varepsilon,1)h_\varepsilon(0,1)}{h_\varepsilon(\varepsilon,1)}. \tag{2.209}$$

The last inequality holds because $w_\varepsilon(0,1) \ge w_\varepsilon(\varepsilon,1)$. We note that for a positive random variable X_1, if $\delta < \langle X_1\rangle/2$ then[33]

$$E[X_1 - \delta; X_1 > \delta] \ge cE[X_1] \quad \text{where } c > 0 \text{ is a universal constant.}$$

We conclude that[34]

$$h_\varepsilon(\varepsilon,1) = \int_\varepsilon^\infty (x-\varepsilon)c_\varepsilon(\varepsilon,1)dx \ge c\int_0^\infty xc_\varepsilon(x,1)dx = ch_\varepsilon(0,1), \tag{2.210}$$

provided $\varepsilon < 1/2$. Since $\beta_\varepsilon(\varepsilon,1) \le 1$, from (2.209), we have an upper bound on $d\Lambda_\varepsilon(t)/dt$ at $t = 1$.

Now we prove the upper bound for $t > 1$. We define a function $\tau(\lambda)$, $\lambda \ge 1$, as the solution to the equation $\Lambda_\varepsilon(\lambda\tau(\lambda)) = \lambda$. It is clear that $\tau(\lambda)$ exists if $\Lambda_\varepsilon(\cdot)$ is a

---

[33] Firstly we have $E[X - \frac{1}{2}\langle X\rangle; X > \frac{1}{2}\langle X\rangle] \ge E[X - \frac{1}{2}\langle X\rangle; X > \frac{3}{4}\langle X\rangle]$. When $X > \frac{3}{4}\langle X\rangle$, $X - \frac{1}{2}\langle X\rangle \ge \frac{X}{3}$, thus $E[X - \frac{1}{2}\langle X\rangle; X > \frac{1}{2}\langle X\rangle] \ge \frac{1}{3}E[X; X > \frac{3}{4}\langle X\rangle]$. Since we also know $E[X] \le E[X; X > \frac{3}{4}\langle X\rangle] + \frac{3}{4}\langle X\rangle \cdot P(X < \frac{3}{4}\langle X\rangle)$, thereby $E[X; X > \frac{3}{4}\langle X\rangle] \ge \frac{1}{4}E[X]$. Therefore $E[X - \frac{1}{2}\langle X\rangle; X > \frac{1}{2}\langle X\rangle] \ge \frac{1}{12}E[X]$.

[34] We have equality

$$\int_\varepsilon^\infty xc_\varepsilon(x,1)dx = \int_\varepsilon^\infty -x\cdot dw_\varepsilon(x,1) = \varepsilon\cdot w_\varepsilon(\varepsilon,1) + \int_\varepsilon^\infty w_\varepsilon(x,1)dx = \varepsilon\int_\varepsilon^\infty c_\varepsilon(x,1)dx - h_\varepsilon(\varepsilon,1).$$

strictly increasing function. To show this, according to the expression of $d\Lambda_\varepsilon(t)/dt$ as in (2.26), we only need to show that $\partial c_\varepsilon(0,t)/\partial x > 0$ for all $t > 0$, which follows from the *Hopf maximum principle*[35]. Furthermore, the function $\tau(\cdot)$ is continuous.

By rescaling, $\lambda^2 c_\varepsilon(\lambda x, \lambda t)$ together with $\Lambda_\varepsilon(\lambda t)/\lambda$ are solutions to (2.22) and (2.23) (An explanation of why we rescale this way will be explained in the remark below). Thus, based on the conclusion from the previous paragraph, we have

$$\frac{d}{dt}\Lambda_\varepsilon(\lambda[\tau(\lambda) + t]) \leq C\lambda \quad \text{at } t = 1. \tag{2.211}$$

Hitherto, we have shown that $d\Lambda_\varepsilon(t)/dt \leq C$ at $t = \lambda[\tau(\lambda) + 1]$. Since the function $\lambda \to \lambda\tau(\lambda)$ is monotonically increasing with range $[0, \infty)$ the result follows. $\square$

Remark II.31. *The functions* $\tilde{c}_\varepsilon(x,t) := \lambda^2 c_\varepsilon(\lambda x, \lambda t)$ *and* $\tilde{\Lambda}_\varepsilon(t) := \Lambda_\varepsilon(\lambda t)/\lambda$ *satisfy the differential equation*

$$\frac{\partial \tilde{c}_\varepsilon(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[1 - \frac{x}{\tilde{\Lambda}_\varepsilon(t)}\right]\tilde{c}_\varepsilon(x,t) + \frac{\varepsilon}{2\lambda}\frac{\partial^2 \tilde{c}_\varepsilon}{\partial x^2}. \tag{2.212}$$

We notice that the diffusive term diminishes as $\lambda \to \infty$.

Proposition II.32. *Suppose* $c_0 : [0, \infty) \to \mathbf{R}^+$ *satisfies* (2.168) *and* $c_\varepsilon(x,t)$, $x \geq 0, t > 0$ *is the solution to* (2.22), (2.23). *Assume that the function* $h_\varepsilon(x,t)$ *is log concave in* x *at* $t = 0$. *Then there exists a universal constant* C, *and* $T \geq 0$ *depending on* $c_0(\cdot), \varepsilon$, *such that* $d\Lambda_\varepsilon(t)/dt \leq C$ *for* $t \geq T$.

Proof. By Lemma II.27 there exists $T_\varepsilon \geq 0$ such that $\varepsilon/\Lambda_\varepsilon(T_\varepsilon) \leq \varepsilon_0$ where ε_0 is the universal constant in Lemma II.30. We do rescaling as in Lemma II.30 with $\lambda = \Lambda_\varepsilon(T_\varepsilon)$. Based on the discussion in II.31, it follows that we can choose $T = T_\varepsilon + \Lambda_\varepsilon(T_\varepsilon)$. $\square$

[35] See Hopf's Lemma [13, page 347].

APPENDIX

APPENDIX A

Supplementary proofs for Chapter I

Proof of Lemma I.8. **Part (*i*).** It suffices to prove $L_d(x)$ and $L_u(x)$ are convex, where

$$\begin{aligned} L_d(x) &= \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi, \\ L_u(x) &= \frac{2}{\sigma_1^2} \int_0^x e^{-\lambda_1(x-\xi)} h(\xi) d\xi + c e^{-\lambda_1 x} \end{aligned}$$

Let $x_1 < x_2$ and $0 \le \alpha \le 1$. We have

$$\begin{aligned} L_d(\alpha x_1 + (1-\alpha)x_2) &= \int_{\alpha x_1 + (1-\alpha)x_2}^\infty e^{-\lambda_0(\xi - \alpha x_1 - (1-\alpha)x_2)} h(\xi) d\xi \\ &= \int_0^\infty e^{-\lambda_0 y} h(y + \alpha x_1 + (1-\alpha)x_2) dy, \end{aligned}$$

and

$$\alpha L_d(x_1) = \alpha \int_{x_1}^\infty e^{-\lambda_0(\xi - x_1)} h(\xi) d\xi = \alpha \int_0^\infty e^{-\lambda_0 y} h(y + x_1) dy,$$

$$(1-\alpha) L_d(x_2) = (1-\alpha) \int_{x_2}^\infty e^{-\lambda_0(\xi - x_2)} h(\xi) d\xi = (1-\alpha) \int_0^\infty e^{-\lambda_0 y} h(y + x_2) dy.$$

Since $h(\cdot)$ is convex, $h(y + \alpha x_1 + (1 - \alpha k)x_2) \le k h(y + x_1) + (1-\alpha) h(y + x_2)$, we obtain

$$L_d(\alpha x_1 + (1-\alpha)x_2) \le \alpha L_d(x_1) + (1-\alpha) L_d(x_2),$$

and the convexity of $L_d(x)$ follows directly.

For $L_u(x)$, since $e^{-\lambda_1 x}$ is convex, it suffices to prove the convexity of the first term. More specifically, we prove that its second order derivative is positive.

$$\begin{aligned}
\frac{d}{dx}\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi &= h(x) - \lambda_1\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi, \\
\frac{d^2}{dx^2}\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi &= h'(x) - \lambda_1 h(x) + \lambda_1^2\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi.
\end{aligned}$$

Because $h(x)$ is convex, for any $\xi \geq 0$ we have $h(\xi) \geq h(x) + h'(x)(\xi - x)$. Thus

$$\begin{aligned}
&\frac{d^2}{dx^2}\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi \\
&\geq h'(x) - \lambda_1 h(x) + \lambda_1^2\int_0^x e^{-\lambda_1(x-\xi)}\left(h(x) + h'(x)(\xi - x)\right)d\xi \\
&= h'(x) - \lambda_1 h(x) + \lambda_1\left(1 - e^{-\lambda_1 x}\right)h(x) - \lambda_1\left(-xe^{-\lambda_1 x} + \frac{1}{\lambda_1} - \frac{1}{\lambda_1}e^{-\lambda_1 x}\right)h'(x) \\
&= \lambda_1 e^{-\lambda_1 x}\Big(-h(x) + xh'(x)\Big) + h'(x)e^{-\lambda_1 x} \\
&\geq \lambda_1\left(-h(0)\right)e^{-\lambda_1 x} + h'(x)e^{-\lambda_1 x} \\
&= h'(x)e^{-\lambda_1 x},
\end{aligned}$$

where the second inequality above follows from, by convexity of $h(\cdot)$, that $h(x) \leq h(0) + xh'(x)$ for all $x \geq 0$, and the last equality follows from $h(0) = 0$. Therefore,

$$\begin{aligned}
&\frac{d^2}{dx^2}\left(\frac{2}{\sigma_1^2}\int_0^x e^{-\lambda_1(x-\xi)}h(\xi)d\xi + ce^{-\lambda_1 x}\right) \\
\geq\ &\left(\frac{2}{\sigma_1^2}h'(x) + c\lambda_1^2\right)e^{-\lambda_1 x} \\
\geq\ &0.
\end{aligned}$$

This proves that $L_u(x)$ is convex. Since the second term of $L_u(x)$, $ce^{-\lambda_1 x}$, is strictly convex, $G(x)$ is also strictly convex. This completes the proof of Part (i).

Part (*ii*). Since $h(x)$ is increasing convex, so $\lim_{x\to\infty} h(x) = \infty$. As $x \to \infty$,

$$\int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi \geq h(x) \int_x^\infty e^{-\lambda_0(\xi - x)} d\xi = \frac{1}{\lambda_0} h(x) \to \infty.$$

This guarantees that $\lim_{x\to\infty} G(x) = \infty$. □

Proof of Proposition I.22. Similar to the proof of Theorem I.7, we compute the expected total cost over a cycle for a band policy with known s, S. Define $w_d(x)$ and $w_u(x)$ respectively as the holding cost incurred during the downward stage from x to s and during the upward stage from x to S, in parallel with those defined in the lost-sales model. Similar to equation (1.13), we obtain the expression of w_d and w_u as

$$\begin{aligned} w_d(x) &= \int_s^x \left(\frac{2}{\sigma_0^2} \int_u^\infty e^{-\lambda_0(\xi - u)} h(\xi) d\xi \right) du \\ w_u(x) &= \int_x^S \left(\frac{2}{\sigma_1^2} \int_{-\infty}^u e^{\lambda_1(\xi - u)} h(\xi) d\xi \right) du. \end{aligned}$$

Therefore, the expected total cost incurred over one cycle is

$$\int_s^S G(x) dx + K.$$

The expected duration of each cycle is $m(S - s)$, with $m = 1/\mu_0 + 1/\mu_1$. The average cost of the band policy (s, S) equals to the ratio of the expected cost to the expected duration, thus (1.38) follows. □

Proof of Lemma I.23. **Part (*i*).** The proof of this part is similar to Part (i) in Lemma I.8. Define

$$\begin{aligned} L_d(x) &:= \int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi, \\ L_u(x) &:= \int_{-\infty}^x e^{-\lambda_1(x - \xi)} h(\xi) d\xi. \end{aligned}$$

Then prove $L_d(x)$ and $L_u(x)$ are convex respectively.

Part (*ii*). As $x \to \infty$,

$$\int_x^\infty e^{-\lambda_0(\xi - x)} h(\xi) d\xi \geq \inf_{y \geq x} h(y) \int_x^\infty e^{-\lambda_0(\xi - x)} d\xi = \frac{1}{\lambda_0} \inf_{y \geq x} h(y) \to \infty.$$

The other part as $x \to -\infty$ is proved similarly.

Part (*iii*). It is easily seen that the second term $K/(S-s)$ is strictly convex. For the first term, we use the approach of Zhang [44], by noting that it can be expressed as $E[G(s + (S - s)U)]$, where U is a continuous uniform random variable on $[0, 1]$. Since for any realization of U, $G(s + (S - s)U)$ is convex in (s, S), it follows that $E[G(s + (S - s)U)]$ is also convex in (s, S). This proves that $c(s, S)$ is strictly convex with respect to s and S. □

Lemma A.1. *If $h(\cdot)$ is polynomially bounded with degree n, then the relative value function v defined in §2 and §3 are polynomially bounded with degree $n + 1$.*

Proof. We only prove the result for the backlog model. By the definition of v, it suffices to prove that $w_d(x), x \geq s$ and $w_u(x), x \leq S$ are both polynomially bounded with degree $n + 1$. We only consider the case that s is non-negative. The case with negative s only adds a constant to the upper bound. We prove by induction that if $h(x) \leq A_1|x|^n$ for some constant A_1, then $w_i(x) \leq A_2 + A_3|x|^{n+1}$, $i = d, u$, for some constants A_2 and A_2. When $n = 0$,

$$\begin{aligned} w_d(x) &\leq \int_s^x \left(\frac{2}{\sigma_0^2} \int_u^\infty e^{-\lambda_0(\xi - u)} A_1 d\xi \right) du \\ &= \frac{2A_1}{\lambda_0 \sigma_0^2}(x - s). \end{aligned}$$

Suppose we have shown that $h(x) \leq A_1|x|^i$ implies $w_d(x) \leq A_2|x|^{i+1}$, for $i =$

$0, 1, \ldots, n-1$, then for n, we have

$$\begin{aligned}
w_d(x) &= \int_s^x \left(\frac{2}{\sigma_0^2} \int_u^\infty e^{-\lambda_0(\xi-u)} h(\xi) d\xi \right) du \\
&\leq \int_s^x \left(\frac{2}{\sigma_0^2} \int_u^\infty e^{-\lambda_0(\xi-u)} A_1 |\xi|^n d\xi \right) du \\
&\leq \int_0^x \left(\frac{2}{\sigma_0^2} \int_u^\infty e^{-\lambda_0(\xi-u)} A_1 \xi^n d\xi \right) du \\
&= -\int_0^x \frac{2A_1}{\sigma_0^2 \lambda_0} \left(e^{-\lambda_0(\xi-u)} \xi^n \Big|_u^\infty - \int_u^\infty n e^{-\lambda_0(\xi-u)} \xi^{n-1} d\xi \right) du \\
&= \frac{2A_1}{\sigma_0^2 \lambda_0} \int_0^x \left(u^n + \int_u^\infty n e^{-\lambda_0(\xi-u)} \xi^{n-1} d\xi \right) du \\
&\leq A_2 |x|^{n+1},
\end{aligned}$$

where the last inequality follows from induction assumption, and A_2 is a constant.

The proof for $w_u(x)$ to be polynomially bounded of degree $n+1$ is similar and is omitted. Thus, the proof of Lemma A.1 is complete. $\square$

Proof of Theorem I.25. We first show that $c(s, S)$ is strictly and jointly convex in (s, S). By substitution $x = uS + (1-u)s$, we can rewrite (1.38) as

$$c(s, S) = \frac{1}{m} \int_0^1 G\left(uS + (1-u)s\right) du + \frac{K}{m(S-s)}.$$

Since $G(x)$ is convex and $uS+(1-u)s$ is an affine function of s and S, $G(uS+(1-u)s)$ is jointly convex in s and S for all $u \in [0, 1]$. Thus $\int_0^1 G(uS+(1-u)s)du$ is also jointly convex in s and S. Further, it is easy to show that $\frac{K}{S-s}$ is strictly jointly convex in s and S on $s < S$. Therefore $c(s, S)$ is a strictly convex function.

The first order optimality condition on $c(s, S)$ with respect to s and S yields

$$G(s) = G(S) = \frac{\int_s^S G(x)dx + K}{S-s}.$$

Let $G(s) = G(S) = m\gamma$, then the optimality condition has three equations

$$G(s) = m\gamma,\ G(s) = m\gamma,$$

$$\int_s^S (G(x) - m\gamma)dx = -K.$$

This establishes (1.39). Since $c(s, S)$ is jointly convex in s and S, the first order necessary optimality condition is also sufficient for optimality. Thus, we only need to prove the existence of (s, S) that satisfy these equations. To that end, define $\ell_\gamma(s, S) = \int_s^S (G(x) - \gamma m)dx + K$ and let $s(\gamma)$ and $S(\gamma)$ be the minimizer of $\ell_\gamma(s, S)$ for given γ, whenever they exist, then $s(\gamma)$ and $S(\gamma)$ are given by (1.18) and (1.19) after replacing $H(\cdot)$ by m, and by continuity of $G(\cdot)$ (since it is convex), we have $G(s(\gamma)) = m\gamma$ and $G(S(\gamma)) = m\gamma$. Furthermore,

$$A(\gamma) = \int_{s(\gamma)}^{S(\gamma)} \Big(G(x) - \gamma m\Big) dx$$

is strictly decreasing and concave in γ. Then, similar argument as that used in the proof of Theorem I.11 shows that, there exists a unique γ^* that satisfies $A(\gamma^*) = -K$. Thus the optimal policy is $s^* = s(\gamma^*)$ and $S^* = S(\gamma^*)$. □

Proof of Theorem I.26. It suffices to prove that the relative value functions defined in (1.40) and (1.41) satisfy (1.34)-(1.37) in Proposition I.21. We firstly prove $\Gamma_0 v(x, 0) + h(x) - \gamma \geq 0$. When $x \geq s^*$, we have

$$\Gamma_0 w_d(x) + h(x) = 0,$$

$$\Gamma_0 \left[-\frac{\gamma^* x}{\mu_0} \right] = \gamma^*,$$

so $\Gamma_0 v(x,0)+h(x)-\gamma=0$. When $x<s^*$, it holds $v(x,0)=v(x,1)+K$. Further,

$$
\begin{aligned}
&\Gamma_0\left[v(x,1)+K\right]\\
=&-\mu_0\left(-\frac{2}{\sigma_1^2}\int_{-\infty}^{x}e^{-\lambda_1(x-\xi)}h(\xi)d\xi+\frac{\gamma^*}{\mu_1}\right)\\
&+\frac{\sigma_0^2}{2}\left[-\frac{2}{\sigma_1^2}\left(h(x)-\lambda_1\int_{-\infty}^{x}e^{-\lambda_1(x-\xi)}h(\xi)d\xi\right)\right].
\end{aligned}
$$

Since

$$
\begin{aligned}
\gamma-h(x) &= \Gamma_0\left[w_d(x)-\frac{\gamma^*x}{\mu_0}\right]\\
&= -\mu_0\left[\frac{2}{\sigma_0^2}\int_{x}^{\infty}e^{-\lambda_0(\xi-x)}h(\xi)d\xi-\frac{\gamma^*}{\mu_0}\right]\\
&\quad+\frac{\sigma_0^2}{2}\left[\frac{2}{\sigma_0^2}\left(-h(x)+\lambda_0\int_{x}^{\infty}e^{-\lambda_0(\xi-x)}h(\xi)d\xi\right)\right]
\end{aligned}
$$

for all x, it suffices to prove

$$\Gamma_0\left[v(x,1)+K\right]-\Gamma_0\left[w_d(x)-\frac{\gamma^*x}{\mu_0}\right]\geq 0.$$

On $x<s^*$, we have $G(x)-\gamma^*m\geq 0$ and $G'(x)\leq 0$, i.e.,

$$\frac{2}{\sigma_1^2}\int_{-\infty}^{x}e^{-\lambda_1(x-\xi)}h(\xi)d\xi+\frac{2}{\sigma_0^2}\int_{x}^{\infty}e^{-\lambda_0(\xi-x)}h(\xi)d\xi-\gamma^*\left(\frac{1}{\mu_0}+\frac{1}{\mu_1}\right)\geq 0,$$

and

$$\frac{2}{\sigma_1^2}\left(h(x)-\lambda_1\int_{-\infty}^{x}e^{-\lambda_1(x-\xi)}h(\xi)d\xi\right)+\frac{2}{\sigma_0^2}\left(-h(x)+\lambda_0\int_{x}^{\infty}e^{-\lambda_0(\xi-x)}h(\xi)d\xi\right)\leq 0,$$

thus it follows that

$$\Gamma_0\left[v(x,1)+K\right]-\Gamma_0\left[w_d(x)-\frac{\gamma^*x}{\mu_0}\right]\geq 0.$$

Therefore, $\Gamma_0 v(x,0) + h(x) - \gamma \geq 0$ is satisfied for all x. Similarly, it can be shown that $\Gamma_1 v(x,1) + h(x) - \gamma \geq 0$ for all x. The other two conditions, (1.36), (1.37), can be proved in the same way as that in the lost-sales case, so they are omitted here. Thus, applying Proposition I.21, we conclude that the proposed policy (s^*, S^*) is optimal among all policies in $\mathcal{A}_v$. $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] J. A. Bather. A continuous time inventory model. *Journal of Applied Probability*, 3(2):538–549, Dec. 1966.

[2] E. Bayraktar and M. Egami. On the one-dimensional optimal switching problem. *Mathematics of Operations Research*, 35(1):140–159, 2010.

[3] J. Carr and O. Penrose. Asymptotic behaviour of solutions to a simplified Lifshitz-Slyozov equation. *Physics D: Nonlinear Phenomena*, 124(1):166–176, 1998.

[4] S. Chen. The infinite horizon periodic review problem with setup costs and capacity constraints: A partial characterization of the optimal policy. *Operations Research*, 52(3):409–421, 05 2004.

[5] S. Chen and M. Lambrecht. X-Y band and modified (s, S) policy. *Operations Research*, 44(6):1013–1019, 11 1996.

[6] J. Conlon. On a diffusive version of the Lifschitz-Slyozov-Wagner equation. *J. Nonlinear Science*, 20:463–521, 2010.

[7] J. Conlon. Bounds on coarsening rates for the Lifschitz-Slyozov-Wagner equation. *Archive for Rational Mechanics and Analysis*, 201(1):343–375, 2011.

[8] G. M. Constantinides and S. F. Richard. Existence of optimal simple policies for discounted-cost inventory and cash management in continuous time. *Operations Research*, 26(4):620–636, 1978.

[9] S. W. Dharmadhikari and K. Joag-dev. *Unimodality, convexity, and applications*. Probability and mathematical statistics. Academic Press, 1988.

[10] B. T. Doshi. Controlled one dimensional diffusions with switching costs–average cost criterion. *Stochastic Processes and their Applications*, 8(2):211–227, 1978.

[11] B. T. Doshi. Two-mode control of Brownian motion with quadratic loss and switching costs. *Stochastic Processes and their Applications*, 6(3):277–289, 1978.

[12] K. Duckworth and M. Zervos. A model for investment decisions with switching costs. *The Annals of Applied Probability*, 11(1):239–260, 2001.

[13] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.

[14] A. Federgruen and P. Zipkin. An inventory model with limited production capacity and uncertain demands. i. the average-cost criterion. *Mathematics of Operations Research*, 11:193–207, May 1986.

[15] Rogers L. C. G. and Williams D. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge University Press, 2000.

[16] G. Gallego. Scheduling the production of several items with random demands in a single facility. *Management Science*, 36(12):1579–1592, 1990.

[17] G. Gallego and A. Scheller-Wolf. Capacitated inventory problems with fixed order costs: Some optimal policy structure. *European Journal of Operational Research*, 126(3):603–613, 2000.

[18] M. K. Ghosh, A. Arapostathis, and S. I. Marcus. Optimal control of switching diffusions with application to flexible manufacturing systems. *SIAM Journal on Control and Optimization*, 31(5):1183–1204, September 1993.

[19] J. M. Harrison. *Brownian Motion and Stochastic Flow Systems*. John Wiley & Sons, Inc, 1985.

[20] J. M. Harrison, T. M. Sellke, and A. J. Taylor. Impulse control of Brownian motion. *Mathematics of Operations Research*, 8(3):454–466, 1983.

[21] J. M. Harrison and A. J. Taylor. Optimal control of a Brownian storage system. *Stochastic Processes and their Applications*, 6(2):179–194, 1978.

[22] A. C. Hax and D. Candea. *Production and inventory management*. Prentice-Hall, 1984.

[23] Dai J. and Yao D. Brownian inventory models with convex holding cost, Part 1: Average-optimal controls. http://www.i-journals.org/ssy/, 2011.

[24] Dai J. and Yao D. Brownian inventory models with convex holding cost, Part 2: Discount-optimal controls. http://www.i-journals.org/ssy/, 2011.

[25] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, New York, 1991.

[26] S. Karlin and H. Taylor. *A Second Course in Stochastic Processes*. New York, 1981.

[27] N. Korevaar. Convex solutions to nonlinear elliptic and parabolic boundary value problems. *Indiana Univ. Math. J.*, 32:603–614, 1983.

[28] B. Meerson. Fluctuations provide strong selection in Ostwald ripening. *Phys. Rev. E*, 60:3072–3075, 1999.

[29] S. Nahmias. *Production and Operations Analysis*. McGraw-Hill/Irwin Series Operations and Decision Sciences. McGraw-Hill, 6th edition, 2008.

[30] M. Ormeci, J. G. Dai, and J. V. Vate. Impulse control of Brownian motion: The constrained average cost case. *Operations Research*, 56(3):618–629, 2008.

[31] H. Pham. On the smooth-fit property for one-dimensional optimal switching problem. In *Séminaire de Probabilités XL*, pages 187–199. 2007.

[32] H. Pham, V. Vath, and X. Zhou. Optimal switching over multiple regimes. *SIAM Journal on Control and Optimization*, 48(4):2217–2253, 2009.

[33] András Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.

[34] M. L. Puterman. A diffusion process model for a storage system. *Logistics*, 1:143–159, 1975.

[35] S. F. Richard. Optimal impulse control of a diffusion process with both fixed and proportional costs of control. *SIAM Journal on Control and Optimization*, 15(1):79–91, 1977.

[36] H. Scarf. The optimality of (s,S) policies in the dynamic inventory problem. In P. Suppes, K. Arrow, and S. Karlin, editors, *Mathematical Methods in the Social Sciences*, pages 196–202. Stanford University Press, Stanford, Stanford, CA, USA., 1960.

[37] D. D. Sheng. A class of problems in the optimal control of diffusions with finitely many controls. In Wendell H. Fleming and Luis G. Gorstiza, editors, *Advances in Filtering and Optimal Stochastic Control*, volume 42 of *Lecture Notes in Control and Information Sciences*, pages 353–375. Springer Berlin Heidelberg, 1982.

[38] A. Sulem. A solvable one-dimensional model of a diffusion inventory system. *Mathematics of Operations Research*, 11(1):125–133, 1986.

[39] V. Vath and H. Pham. Explicit solution to an optimal switching problem in the two-regime case. *SIAM Journal on Control and Optimization*, 46(2):395–426, April 2007.

[40] A. F. Veinott, Jr. On the optimality of (s, S) inventory policies: New conditions and a new proof. *SIAM Journal on Applied Mathematics*, 14(5):1067–1083, 09 1966.

[41] R. G. Vickson. Optimal control of production sequences: A continuous parameter analysis. *Operations Research*, 30(4):659–679, 1982.

[42] R. G. Vickson. A single product cycling problem under Brownian motion demand. *Management Science*, 32:1336–1345, October 1986.

[43] J. Wijnaard. An inventory problem with constrained an inventory problem with constrained order capacity. Th-report 72-wsk-63, Eindhoven University of Technology, Eindhoven, The Netherlands, 1972.

[44] H. Zhang. A note on the convexity of service-level measures of the (r,q) system. *Management Science*, 44:431–432, 1998.

[45] P. Zipkin. Inventory service-level measures: convexity and approximation. *Management Science*, 32:975–981, 1986.