CHARACTERIZATION OF OPTIMAL STRATEGY FOR MULTI-ASSET INVESTMENT AND CONSUMPTION WITH TRANSACTION COSTS

XINFU CHEN    MIN DAI

Abstract: We consider the optimal consumption and investment with transaction costs on multiple assets, where the prices of risky assets jointly follow a multi-dimensional geometric Brownian motion. We characterize the optimal investment strategy and in particular prove by rigorous mathematical analysis that the trading region has the shape that is very much needed for well defining the trading strategy, e.g., the no-trading region has distinct corners. In contrast, the existing literature is restricted to either single risky asset or multiple uncorrelated risky assets.

Keywords: portfolio selection, optimal investment and consumption, transaction costs, multiple risky assets, shape of trading and no-trading regions.

1. INTRODUCTION

We consider the optimal investment and consumption decision of a risk-averse investor who has access to multiple risky assets as well as a riskfree asset. Proportional transaction costs are incurred when the investor buys or sells the risky assets whose prices are assumed to follow a multi-dimensional geometric Brownian motion. We aim to provide a theoretical characterization of the optimal strategy.

In the absence of transaction costs, the problem described above has been studied by Merton (1969, 1971). It turns out that the optimal strategy of a constant relative risk aversion (CRRA) investor is to keep a constant fraction of total wealth in each assets and consume at a constant fraction of total wealth. In contrast, the optimal strategy of a constant absolute risk aversion (CARA) investor is to keep a certain fixed amount in each risky asset and a consumption that is affine in the total wealth. Merton’s strategy requires continuous trading in all risky assets and thus must be suboptimal when transaction costs are incurred.

Magill and Constantinides (1976) introduce proportional transaction costs to Merton’s model with single risky asset and a CRRA investor. They provide a fundamental insight that there exists an interval, known as the no-trading region, such that the optimal investment strategy is to keep the fraction of wealth invested in the risky asset within the interval (i.e., no-trading region). Hence, as long as the initial fraction falls within the no-trading region, the future transactions only occurs at the boundary of the region. For a CARA investor, it can be shown that the optimal investment strategy is to keep the dollar amount in the risky asset between two levels [cf. Liu (2004) and Chen et al. (2012)].

Date: First version: March 2012; This version: November 2012.

Xinfu Chen acknowledges support from NSF grant DMS-1008905. Min Dai acknowledges support from Singapore AcRF grant R-146-000-138-112 and NUS Global Asia Institute - LCF Fund R-146-000-160-646.

Department of Mathematics, University of Pittsburgh, PA 15260, USA.

Department of Mathematics and Centre for Quantitative Finance, National University of Singapore, Singapore. Email: matdm@nus.edu.sg.

Most of existing theoretical characterizations of optimal strategy are for the single risky-asset case.1 In contrast, there is relatively limited literature on the multiple risky-asset case. Assuming that there are multiple uncorrelated risky assets available for investment, Akian et al. (1996) obtain some qualitative results on the optimal strategy of a CRRA investor. Liu (2004) considers a CARA investor who is also restricted to invest in uncorrelated risky assets. He shows that the problem can be reduced, by virtue of the separability of the CARA utility function, to the single risky-asset case. This leads to the separability of the optimal investment strategy which is to keep the dollar amount invested in each asset between two constant levels. Unfortunately, such a reduction does not work when the risky assets are correlated.

The main contribution of this paper is to provide a thorough characterization of the optimal investment strategy for a risk-averse investor who can access multiple correlated risky assets as well as a risk-free asset. We focus on the CARA utility case, and an extension to the CRRA utility case is placed in Appendix. To illustrate our results, we take as an example the scenario of two risky assets. We will show that the shape of trading and no-trading regions must be as in Figure 1, where “S_i”, “B_i”, and “N_i” represent selling, buying, and no trading in asset i, respectively. The no-trading region N_1 ∩ N_2 locates in the center, surrounded by eight trading regions. Moreover, each intersection ∂S_1 ∩ ∂S_2, ∂S_1 ∩ ∂B_2, ∂B_1 ∩ ∂B_2, ∂B_1 ∩ ∂S_2, and ∂B_1 ∩ ∂B_2 is a singleton. In addition, we show that the boundary of each of corner regions S_1 ∩ S_2, S_1 ∩ B_2, B_1 ∩ B_2, and B_1 ∩ S_2 consists of one vertical and one horizontal half line, whereas the boundary of each of S_1 ∩ N_2, N_1 ∩ S_2, B_1 ∩ N_2, and N_1 ∩ B_2 consists of two parallel either vertical or horizontal half lines and a curve in between connecting the end points of the two half lines. These characterizations on the

---

1There do exist many papers working on perturbation analysis or numerical solutions for the multiple risky-asset case, e.g. Law et al. (2009), Bichuch and Shreve (2011), Muthuraman and Kumar (2006), Dai and Zhong (2010).
shapes of trading regions are extremely important because they are necessary conditions for the trading strategy to be well-defined.\textsuperscript{2} For example, given an initial portfolio in $S_1 \cap B_2$, the investor should sell asset 1 and buy asset 2 to the unique corner $\partial S_1 \cap \partial B_2$; similarly, given an initial portfolio in $S_1 \cap N_2$, the investor should sell asset 1 and keep asset 2 unchanged to the unique portfolio on $\partial S_1 \cap N_2$.

We also prove that the no-trading region is contained in a union of uniformly bounded ellipses. Thus in numerical simulation one need only perform computation on the bounded union. Furthermore, we provide a precise characterization of the corners of the no-trading region.

Since negative wealth is permitted with the CARA utility, we need to impose some constraint to prevent the investor from unlimited consumptions. This motivates us to start from a finite horizon problem formulation (Section 2), where the expected utility is from not only the intermediate consumption but also the terminal wealth (i.e., bequest). Section 3 is devoted to some basic properties of the value function associated with the finite horizon problem. In Section 4 we derive the infinite horizon problem as the limit of the finite horizon problem. In Section 5, we show the $C^1$ continuity of the resulting value function that is useful for analyzing the optimal strategy. The main results are presented in section 6.

\section{Problem Formulation}

Consider a portfolio consisting of one risk-free asset (bank account) and $n$ risky assets whose unit share prices are stochastic process $\{ (S^0_t, S^1_t, \cdots, S^n_t) \}$ described by the stochastic differential equations

$$\frac{dS^0_t}{S^0_t} = rd\!t, \quad \frac{dS^i_t}{S^i_t} = \alpha_i dt + \sum_{j=1}^{n} a_{ij} dW^j_t$$ \text{ for } i = 1, \cdots, n,

\textsuperscript{2}It should be pointed out that these results have been conjectured or numerically verified by some researchers [e.g. Liu (2004), Dai and Zhong (2010), Bichuch and Shreve (2011)]. However no theoretical analysis has been given so far.
where \( \{W^i_t, \cdots, W^n_t\}_{t \geq 0} \) is a standard n-dimensional Wiener process, \( r > 0 \) is the constant bank rate, \( \alpha_i > 0 \) is the expected return rates of the \( i \)-th risky asset, and \((a_{ij})_{n \times n}\) is a positive definite matrix. We consider optimal strategies of investment and consumption subject to transaction cost which are proportional to the amount of transactions.

2.1. Investment and Consumption.

We introduce a non-negative parameter \( \kappa \) where \( \kappa = 0 \) corresponds to the no-consumption case. Suppose the terminal time is \( T \) and current time is \( t < T \). For \( s \in [t, T) \), we denote by \( \kappa c_s ds \) the consumption, deducted from the bank account, during time interval \([s, s + ds]\).

Here we assume that \( \kappa \) has the same unit as \( r \), being 1/year, and that \( c_s \) has the unit of dollars.\(^3\) We denote by \( dL^i_s \) the transfer of money from the bank account to the \( i \)-th risky assets during \([s, s + ds]\), which incurs purchasing costs \( \lambda dL^i_s \). Similarly, we denote by \( dM^i_s \) the money transferred from the \( i \)-th risky asset to the bank account during \([s, s + ds]\), which incurs selling costs \( \mu_i dM^i_s \). Here \( \lambda_i \geq 0 \) and \( \mu_i \in [0, 1) \) are the constant proportions of transaction costs for purchasing and selling the \( i \)-th risky asset, respectively.

Let \( x_s \) and \( y_s = (y^1_s, \cdots, y^n_s) \) be dollar values at time \( s \in [t, T] \) invested in the bank account and risky assets, respectively. Their evolutions are described by

\[
\begin{cases}
  dx_s = (rx_s - \kappa c_s)ds - \sum_i (1 + \lambda_i)dL^i_s + \sum_i (1 - \mu_i)dM^i_s, \\
  dy^i_s = y^i_s(\alpha^i ds + \sum_j a^i_{ij} dW^j_s) + dL^i_s - dM^i_s, \quad i = 1, \cdots, n.
\end{cases}
\tag{2.1}
\]

For simplicity, we define an admissible (investment-consumption) strategy as \( S = (C, L, M) \) where \( C = \{c_s\}_{s \in [t,T]} \), \( L = \{L^1_s, \cdots, L^n_s\}_{s \in [t,T]} \), and \( M = \{M^1_s, \cdots, M^n_s\}_{s \in [t,T]} \) are adapted processes satisfying

\[
  dL^i_s \geq 0, \quad dM^i_s \geq 0, \quad \sup_{t \leq s \leq T} \left( \|c_s\|_{L^\infty} + \|x_s\|_{L^\infty} + \|y^i_s\|_{L^\infty} \right) < \infty.
\tag{2.2}
\]

where \( \{y^i_s\}_{t \leq s \leq T} \) is the solution of the second set of equations in (2.1) subject to constant initial conditions. We denote by \( \mathcal{A} \) all the admissible strategies. Here we remark that the “optimal strategy” may not be attained, but can be approximated, by the admissible strategies constrained by (2.2).\(^4\)

2.2. The Merton’s Problem.

Given concave utilities \( U(x, y) \) for the terminal portfolio, \( V_c \) for consumption, a discount factor constant \( \beta > 0 \), and positive dimensionless constant weight \( K \), we consider the measure of quality of an investment-consumption strategy \( S \) defined by

\[
J(S, t) := KU(x_T, y_T)e^{-\beta(T-t)} + \int_t^T V(c_s)e^{-\beta(s-t)}Kds,
\tag{2.3}
\]

\(^3\)The parameters \( \kappa \) and \( K \) in (2.3) below are both used as the weight between consumption and terminal wealth. However \( K \) is dimensionless while \( \kappa \) has the same unit as \( r \). It should be pointed out that the condition \( r < 1 \) in Chen et al. (2012) should be replaced by \( r < \kappa \). If the issue of unit for \( \kappa \) and \( K \) is neglected, we can simply write \( K = 1 - \kappa \).

\(^4\)For a given type of investor, we may expand the set of admissible strategies such that the “optimal strategy” is admissible. However, as mentioned later, constructing the “optimal strategy” requires certain regularity of the boundary of the no-trading region, which is not covered by the present paper.
where \( \{x_s, y_s\}_{s \in [t,T]} \) is the solution of (2.1) with given strategy \( S \in \mathcal{A}^t \). The Merton’s problem is to maximize the expected utility:

\[
\Phi(x, y, t) = \sup_{S \in \mathcal{A}_t} \mathbb{E}^x_{t} [J(S, t)] \quad \forall x \in \mathbb{R}, y \in \mathbb{R} \times \mathbb{R}^n, t \leq T,
\]

where \( \mathbb{E}^x_{t} \) is the expectation under the condition \((x_t, y_t) = (x, y)\). In this paper, we consider the exponential utility

\[
V(c) := -e^{-\gamma c}, \quad U(x, y) := V(x + \ell(y)),
\]

where \( \ell(y) \) is the liquidation value of the holdings in the risky assets:

\[
\ell(y) = \sum_i \ell_i(y_i), \quad \ell_i(y_i) = \begin{cases} (1 - \mu_i)y_i, & \text{if } y_i \geq 0, \\ (1 + \lambda_i)y_i, & \text{if } y_i < 0, \end{cases}
\]

Notice that \( \ell_i(\cdot) \) is a concave function and

\[
\ell_i(y_i) = \min_{1 - \mu_i \leq \xi \leq 1 + \lambda_i} \{ k y_i \} = \min\{(1 - \mu_i)y_i, (1 + \lambda_i)y_i\} \quad \forall y_i \in \mathbb{R}.
\]

### 3. Basic Properties of \( \Phi \)

#### 3.1. The Case of No Risky Assets.

Here we establish a useful lower bound of \( \Phi(x, 0, t) \) by considering the category of strategies that do not use risky assets; that is, we consider the strategies where \( y_s = 0, L_s = 0, M_s = 0 \) for all \( s \in [t,T] \). Writing \((c_o, x_s)\) as \((c(s), x(s))\), we have \( dx(s) = [r x(s) - \kappa c(s)] ds \).

Subject to \( x(t) = x \) we obtain

\[
x(T) = xe^{r(T-t)} - \int_t^T e^{r(T-s)} c(s) ds.
\]

Thus, for any consumption strategy \( c \in L^\infty \), the total utility can be written as

\[
J_0^{x, t}[c] := -\int_t^T e^{-\beta(s-t) - \gamma c(s)} ds - K e^{-\gamma x e^{rt}(T-t)} - \gamma \int_t^T e^{\gamma c(s) ds} - \beta(T-t).
\]

We want to find an optimal consumption strategy that maximizes \( J_0^{x, t} \).

When \( \kappa = 0 \), we have \( J_0^{x, t}[c] = -K e^{-\gamma x e^{rt}(T-t)} - \beta(T-t) \).

Next consider the case \( \kappa > 0 \). The first variation of \( J_0^{x, t} \) can be calculated by

\[
\left\langle \frac{\delta J_0^{x, t}[c]}{\delta c}, \zeta \right\rangle := \lim_{h \to 0} \frac{J_0^{x, t}[c + h \zeta] - J_0^{x, t}[c]}{h} = \kappa \gamma \int_t^T e^{-\beta(s-t) - \gamma c(s)} \left\{ e^{-\gamma c(s)} - K e^{-\gamma x e^{rt}(T-t)} \right\} ds,
\]

where \( x(T) \) is as given in (3.1). Hence, if \( c^* \) is a critical point of \( J_0^{x, t} \), i.e., \( \delta J_0^{x, t}[c^*] / \delta c = 0 \), then \( e^{-\gamma c(s)} - K e^{-\gamma x e^{rt}(T-t)} = 0 \), where \( x^t(T) \) is as in (3.1) with \( c \) replaced by \( c^* \). Thus \( c^t(s) = x^t(T) - [(r - \beta)(T-t) + \ln K] / \gamma \). Using the definition of \( x^t(T) \) we then obtain

\[
c^t(s) = x \xi(\tau) + \frac{(r - \beta)[s - t - b(\tau)] - Z(\tau) \ln K}{\gamma} \quad \forall s \in [t,T],
\]

where \( \tau = T - t \) and

\[
\xi(\tau) = \frac{re^{r\tau}}{r + \kappa e^{r\tau} - \kappa}, \quad Z(\tau) = e^{-r\tau}, \quad b(\tau) = \frac{\kappa(e^{r\tau} - 1 - r\tau) + \tau^2}{r \left( r + \kappa e^{r\tau} - \kappa \right)}.
\]

(4.4)
It is easy to verify that
\begin{align}
\xi' &= (r - \kappa \xi)\xi \quad \text{on } [0, \infty), \quad \xi(0) = 1; \\
Z' &= -\kappa \xi Z \quad \text{on } [0, \infty), \quad Z(0) = 1; \\
b' &= -\kappa \xi b + 1 \quad \text{on } [0, \infty), \quad b(0) = 0.
\end{align}

Note that $J_0^{x,t}$ is a concave functional, so $c^*$ is the global maximizer.

We have proved the following result:

**Lemma 3.1.** For each $x \in \mathbb{R}$, the linear function $c^*$ defined in (3.3), where $\tau = T - t$ and $\xi, Z, \text{and } b$ are as in (3.4), is the global maximizer of $J_0^{x,t}$ defined in (3.2):

$$J_0^{x,t}[c] \leq J_0^{x,t}[c^*] \quad \forall c \in \mathcal{L}^\infty.$$

The strategy that liquidates all risky assets at time $t$ gives the estimate

$$\Phi(x, y, t) \geq \Phi(x + \ell(y), 0, t) \geq J_0^{x+\ell(y), t}[c^*] = -e^{-\gamma \tau(x + \ell(y)) + (r - \beta)b(T) - \ln \xi(T) + Z(T) \ln K}.$$

Then we have the following corollary:

**Corollary 3.2.** We have the following lower bound for $\Phi$:

$$\Phi(x, y, t) \geq -e^{-\gamma \tau(x + \ell(y)) + (r - \beta)b(T) - \ln \xi(T) + Z(T) \ln K}.$$

### 3.2. Separation of Investment and Consumption.

**Lemma 3.3.** Let $\tau = T - t$ and $\xi$ be as defined in (3.4). Then

$$\Phi(x, y, t) = e^{-\gamma \xi(T)x} \Phi(0, y, t) \quad \forall (x, y, t) \in \mathbb{R} \times \mathbb{R}^n \times (-\infty, T].$$

**Proof.** Let $S = (C, L, M)$ be an investment-consumption strategy for initial position $(x_t, y_t) = (0, y)$, resulting in the subsequent portfolio $\{(x_s, y_s)\}_{s \in [t, T]}$. For another initial position $(x, y)$ at time $t$, we consider the investment-consumption strategy $\tilde{S} = (\tilde{C}, \tilde{L}, \tilde{M})$ defined by

$$(\tilde{L}, \tilde{M}) \equiv (L, M), \quad \tilde{c}_s = c_s + \xi(T)x, \quad \tau := T - t.$$

Denote the corresponding portfolio (starting from $(x, y)$ at time $t$) by $\{(\tilde{x}_s, \tilde{y}_s)\}_{s \in [t, T]}$. Then $\tilde{y}_s = y_s$ and $\tilde{x}_s = x_s + \tilde{x}(s)$ where $\tilde{x}(s)$ is the solution of $d\tilde{x}(s) = [r\tilde{x}(s) - \kappa \xi(T)x]ds$ subject to $\tilde{x}(t) = x$. Solving this initial value problem for $\tilde{x}$ gives

$$\tilde{x}(T) = \frac{\kappa \xi(T)x}{r} + \left( x - \frac{\kappa \xi(T)x}{r} \right) e^{-\gamma \tau} = \xi(T)x$$

by the definition of $\xi(T)$ in (3.4). It then follows that

$$J(\tilde{S}, t) = -Ke^{-\gamma \tilde{x}(T)U(x_T, y_T)} e^{-\beta(T - t)} - \int_t^T e^{-\gamma \xi(T)x} e^{-\gamma c_s - \beta(T - s)K} ds = e^{-\gamma \xi(T)x} J(S, t).$$

The relation between $S$ and $\tilde{S}$ is 1-1 and onto, so taking the supremum yields (3.9). \qed
3.3. Super-Solution.

Let $\varphi(x, y, t)$ be a smooth function in $\mathbb{R} \times \mathbb{R}^n \times (-\infty, T]$ satisfying $\partial_x \varphi > 0$. Let $\mathcal{S}$ be an investment-consumption strategy in $A^t$. By Itô’s formula and (2.1),

$$
\varphi(x_t, y_t, t) = \varphi(x_T, y_T, T)e^{-\beta(T-t)} - \int_t^T \mathbb{E}^x_y \left[ \varphi(x_s, y_s, s) e^{-\beta(s-t)} \right] ds - \int_t^T e^{-\beta(s-t)} \sum_{i,j} a_{ij} y_i \partial_{y_j} \varphi dB^i_s
$$

where $L^i$ and $M^i$ are the continuous part of $L^i$ and $M^i$ respectively,

$$
V^*(q) := \max_{c \in \mathbb{R}} \left\{ V(c) - cq \right\} \quad \forall q > 0,
$$

$$
\mathcal{L} \varphi := \frac{1}{2} \sum_{i,j} \sigma_{ij} y_i y_j \partial_{y_i} \varphi + \sum_i \alpha_i y_i \partial_{y_i} \varphi + r x \partial_x \varphi - \beta \varphi + \kappa V^* \left( \partial_x \varphi \right), \quad (3.10)
$$

Now suppose $\varphi$ satisfies $\varphi(\cdot, T) \geq KU(\cdot)$, $-\partial_x \varphi - \mathcal{L} \varphi \geq 0$, and

$$
(1 + \lambda_i) \partial_x \varphi - \partial_{y_i} \varphi \geq 0, \quad (-1 + \mu_i) \partial_x \varphi + \partial_{y_i} \varphi \geq 0. \quad (3.11)
$$

Combination of (3.11) and $\partial_x \varphi > 0$ leads to

$$
\varphi(x_{s-}, y_{s-}, s-) - \varphi(x_s, y_s, s) \geq 0.
$$

Taking the expectation we obtain

$$
\varphi(x, y, t) \geq \mathbb{E}^x_y \left[ KU(x_T, y_T) e^{-\beta(T-t)} + \int_t^T V(c_s) e^{-\beta(s-t)} (kd\tau) \right].
$$

Taking the supremum we obtain the following:

**Lemma 3.4.** Suppose $\varphi$ is a smooth function on $\mathbb{R} \times \mathbb{R}^n \times (-\infty, T]$ satisfying $\varphi_x > 0$, $\varphi(\cdot, T) \geq KU(\cdot)$ and (3.11). Then $\Phi(x, y, t) \leq \varphi(x, y, t)$ for all $(x, y, t) \in \mathbb{R} \times \mathbb{R}^n \times [0, T]$.

We call such $\varphi$ a super-solution. Due to the separation property (3.9), we seek a supersolution of the form

$$
\varphi(x, y, t) = -e^{-\gamma \xi(x) + (r - \beta)b(\tau - t)} - \ln \xi(\tau) + Z(\tau) \ln K - \phi(z, \tau), \quad \tau = T - t, \quad z = \gamma \xi(\tau) y,
$$
where $\xi, Z$ and $b$ are defined in (3.4). We can compute

\[
-\partial_t \varphi - L \varphi = |\varphi| \left\{ \gamma (\xi' - r \xi + \kappa \xi^2) + \left[ \frac{\xi'}{\xi} + \kappa \xi - \beta - (r - \beta) (b' + \kappa b) \right] \right.
- (Z' + \kappa Z) \ln K + \frac{\xi'}{\xi} \sum_i z_i \partial_{z_i} \phi \\
+ \partial_r \phi - \frac{1}{2} \sum_{i,j} \sigma_{ij} z_i z_j (\partial_{z_i z_j} \phi - \partial_{z_i} \phi \partial_{z_j} \phi) - \sum_i \alpha^i z_i \partial_{z_i} \phi + \kappa \xi \phi \left. \right\}
\]

where we have used (3.5)-(3.7) in the last equality, and

\[
A^T[\phi] := \frac{1}{2} \sum_{i,j} \sigma_{ij} z_i z_j \left[ \partial_{z_i z_j} \phi - \partial_{z_i} \phi \partial_{z_j} \phi \right] + \sum_i (\alpha^i - r) z_i \partial_{z_i} \phi + \kappa \xi \left[ z \cdot \nabla \phi - \phi \right] \tag{3.12}
\]

where the dependence of $A^T$ on $\tau$ is via $\xi = \xi(\tau)$. Define

\[
B(p) := \min_i \min \{ 1 + \lambda_i - p_i, -1 + \mu_i + p_i \} \quad \forall p = (p_1, \cdots, p_n) \in \mathbb{R}^n. \tag{3.13}
\]

Then (3.11) can be written as

\[
\min \left\{ \partial_r \phi - A^T[\phi], B(\nabla \phi) \right\} \geq 0 \quad \text{in} \quad \mathbb{R}^n \times (0, \infty).
\]

Since $\xi(0) = 1 = Z(0)$ and $b(0) = 0$, the condition $\phi(\cdot, T) \geq KU(\cdot)$ can be written as

\[
\phi(z, 0) \geq \ell(z) = \sum_i \ell_i(z_i),
\]

where $\ell$ and $\ell_i$ are as defined in (2.4).

### 3.4. Upper Bound

Let $k = (k_0, k_1, \cdots, k_n)$ be a constant vector satisfying

\[
k_0 \geq 0, \quad 1 - \mu_i \leq k_i \leq 1 + \lambda_i \quad \forall i.
\]

Consider the function

\[
\bar{\phi}(k; z, \tau) := k_0 b(\tau) + \sum_i k_i z_i.
\]

It is easy to see that $\bar{\phi}(k; z, 0) = \sum_i k_i z_i \geq \phi_0(z), 1 + \lambda_i - \bar{\partial}_z \phi = 1 + \lambda_i - k_i \geq 0$ and $-1 + \mu_i + \bar{\partial}_z \phi = -1 + \mu_i + k_i \geq 0$ for each $i$. Also,

\[
\bar{\phi}_r - A^T[\phi] = \left( b' + k \xi \right) b(\tau) + \frac{1}{2} \sum_{i,j} (z_i k_i) \sigma_{ij} (z_j k_j) - \sum_i (\alpha_i - r)(z_i k_i)
\]

\[
= k_0 + \frac{1}{2} \sum_{i,j} (z_i k_i - m_i) \sigma_{ij} (z_j k_j - m_j) - A_0,
\]

where, denoting by $(\sigma^{ij})_{n \times n}$ the inverse matrix of $(\sigma_{ij})_{n \times n},$

\[
m_j := \sum_i \sigma^{ji} (\alpha_i - r) \quad \forall j, \quad A_0 := \frac{1}{2} \sum_{i,j} m_i \sigma_{ij} m_j. \tag{3.14}
\]
Here \( \mathbf{m} := (m_1, \ldots, m_n) \) is the optimal strategy for the Merton’s problem without transaction costs, being the solution of the linear system

\[
\sum_j \sigma_{ij} m_j = \alpha_i - r \quad \forall i = 1, \ldots, n.
\]

Since \((\sigma_{ij})_{n \times n}\) is positive-definite, taking \(k_0 = A_0\) we have \(\bar{\phi} - L^T \bar{\phi} \geq 0\). Hence, by Lemma 3.4,

\[
\Phi(x, y, t) \leq -e^{-\gamma \xi(t)x + (r - \beta)b(\tau) + Z(\tau) \ln K - \ln \xi(\tau)} - \bar{\phi}(k, z, \tau).
\] (3.15)

We are now ready to show the following:

**Theorem 1.** There exists a function \( \psi \) defined on \( \mathbb{R}^n \times [0, \infty) \) such that

\[
\Phi(x, y, t) = -e^{-\gamma \xi(t)x + (r - \beta)b(\tau) + Z(\tau) \ln K - \ln \xi(\tau)} - \psi(\gamma \xi(t)y, \tau),
\] (3.16)

where \( \tau = T - t \) and \( \xi, Z, b \) are defined in (3.4). In addition,

1. for each \( \tau \geq 0 \), \( \psi(\cdot, \tau) \) is Lipschitz continuous:

\[
\ell(z - \hat{z}) \leq \psi(z, \tau) - \psi(\hat{z}, \tau) \leq -\ell(z - \hat{z}) \quad \forall z, \hat{z} \in \mathbb{R}^n;
\] (3.17)

2. for \( A_0 \) defined in (3.14) with \( (\sigma_{ij})_{n \times n} \) being the inverse of \( (\sigma_{ij})_{n \times n} \),

\[
\ell(z) \leq \psi(z, \tau) \leq \ell(z) + A_0 b(\tau) \quad \forall z \in \mathbb{R}^n, \tau \geq 0;
\] (3.18)

3. for each \( \tau \geq 0 \), \( \psi(\cdot, \tau) \) is concave;

4. \( \psi \) is a viscosity solution of

\[
\begin{cases}
\min\left\{ \partial_\tau \psi - \mathcal{A}^\tau[\psi], \ B(\nabla \psi) \right\} = 0 & \text{in} \quad \mathbb{R}^n \times (0, \infty), \\
\psi(\cdot, 0) = \ell(z) & \text{on} \quad \mathbb{R}^n \times \{0\},
\end{cases}
\] (3.19)

where \( \mathcal{A}^\tau \) and \( B \) are defined in (3.12) and (3.13).

**Proof.** By (3.8), \( \Phi(x, y, t) > -\infty \). Also, from (3.15) we see that \( \Phi(x, y, t) < 0 \) for every \( (x, y, t) \in \mathbb{R} \times \mathbb{R}^n \times (-\infty, T] \). Hence, we can define

\[
\psi(z, \tau) = (r - \beta)b(\tau) + Z(\tau) \ln K - \ln \xi(\tau) - \ln \left| \Phi(0, \gamma \xi(t)z, T - \tau) \right|.
\] (3.20)

Then by Lemma 3.3, we obtain (3.16). Now we establish properties of \( \psi \).

1. Let \( x \in \mathbb{R}, y \in \mathbb{R}^n \) and \( \hat{y} \in \mathbb{R}^n \). Stating at position \((x, \hat{y})\) at time \( t \), one can immediately liquidate \( \hat{y} - y \) amount of risky asset holding to reach the position \((x + \ell(\hat{y} - y), y)\) at time \( t+ \). Hence, we have

\[
\Phi(x, \hat{y}, t) \geq \Phi(x + \ell(\hat{y} - y), y, t) = e^{-\ell(\gamma \xi(t)\hat{y} - y)} \Phi(x, y, t).
\]

In terms of (3.16), this implies that \( \psi(\hat{z}, \tau) \geq \psi(z, \tau) + \ell(z - \hat{z}) \). Thus, we obtain (3.17).

2. Combination of the estimate (3.8) (with \( y = z/\gamma \xi(\tau) \)) and (3.16) yields the left hand side inequality of (3.18). To show the right hand side inequality, we use (3.15) with \( k_0 := A_0 \) and (3.16) to derive that

\[
\psi(z, \tau) \leq \phi(k, z, \tau) = A_0 b(\tau) + \sum_i k_i z_i \quad \text{whenever} \quad k_i \in [1 - \mu, 1 + \lambda]
\]

for all \( i \). Hence,

\[
\psi(z, \tau) \leq \min_{1 - \mu \leq k_i \leq 1 + \lambda} \left( A_0 b(\tau) + \sum_i k_i z_i \right) = A_0 b(\tau) + \ell(z).
\]
Thanks to the linearity of transaction costs and the concavity of the exponential utility function, we immediately obtain the concaveness of \( \Phi(\cdot, \cdot, t) \). The concaveness of \( \psi(\cdot, \tau) \) follows by noting (3.20) and the fact that the function \( \ln(\cdot) \) is concave and increasing.

That \( \psi \) is a viscosity solution of (3.19) follows by a dynamical programming principle [see, for example, Shreve and Soner (1994)]. This completes the proof of Theorem 1. □

4. The Asymptotic Behavior as \( T \to \infty \)

Since we are interested in the infinite horizon (i.e., \( T = \infty \)) problem, let us consider the asymptotic behavior of \( \psi(\cdot, \tau) \) as \( \tau = T - t \to \infty \). From now on we always assume \( \kappa > 0 \).

For any function \( f \) defined on \( \mathbb{R}^n \), we define its super-differential by

\[
\partial f(z) = \{ p \in \mathbb{R}^n : f(\hat{z}) \leq f(z) + p \cdot (\hat{z} - z) \quad \forall \hat{z} \in \mathbb{R}^n \}.
\] (4.1)

We shall use the following fact.

**Lemma 4.1.** Suppose \( f \) is a concave function on \( \mathbb{R}^n \). Define its super-differential by (4.1).

Then the following holds:

1. The set \( \{ (z, p) \mid z \in \mathbb{R}^n, p \in \partial f(z) \} \) is closed; i.e. if \( p_k \in \partial f(z_k) \) for all \( k \geq 1 \) and \( \lim_{k \to \infty} (p_k, z_k) = (p, z) \), then \( p \in \partial f(z) \).
2. For each \( z \in \mathbb{R}^n \), \( \partial f(z) \) is a non-empty, convex and compact set.
3. If \( \partial f(z) = \{ p \} \) is a singleton, then \( f \) is differentiable at \( z \) and \( p = \nabla f(z) \).
4. If \( \partial f(z) \) is singleton for every \( z \) in an open neighborhood of \( z^0 \in \mathbb{R}^n \), then \( f \) is \( C^1 \) in an open neighborhood of \( z^0 \).
5. For each \( i = 1, \ldots, n \) and fixed \( z \in \mathbb{R}^n \) define

\[
\partial_i f(z) = \{ e_i \cdot p : p \in \partial f(z) \}, \quad g(t) = f(z + te_i).
\]

Then

\[
\partial_i f(z) = \partial g(0) = \left[ \lim_{h \to 0} \frac{g(h) - g(0)}{h}, \lim_{h \to 0} \frac{g(0) - g(-h)}{h} \right].
\]

Most of the conclusion of the Lemma should be well-known. For completeness, we provide the full proof in Appendix.

If \( f \) is concave, then \( f \) is locally Lipschitz continuous and \( \partial f \) is non-empty and almost everywhere singleton, and coincides with the Sobolev gradient. For convenience, we identify the set \( \partial f(z) \) as a generic vector \( p \) in \( \partial f(z) \).

We begin with the following estimate:

**Lemma 4.2.** For any \( z \in \mathbb{R}^n \) and \( \tau \geq 0 \),

\[
0 \leq \psi(z, \tau) - z \cdot \partial \psi(z, \tau) \leq A_0 b(\tau). \tag{4.2}
\]

**Proof.** The assertion holds for \( z = 0 \) by (3.18).

Fix \( z \in \mathbb{R}^n \setminus \{0\} \). Consider the function

\[
f(s) := \psi(sz, \tau), \quad s \in [0, \infty).
\]
This is a concave and Lipschitz continuous function in one space dimension. Hence, \( f'(s) \) is a decreasing function. Set
\[
p_\infty = \lim_{s \to \infty} \frac{f(s) - f(0)}{s}.
\]
In view of (3.17) and the homogeneity \( \ell(sz) = s\ell(z) \) for \( s > 0 \), we find that \( p_\infty \geq \ell(z) \). As \( f'(s) \) is a decreasing function, by L’Hôpital’s rule, \( f'(s) \downarrow p_\infty \) as \( s \to \infty \). Consequently, for any \( s > 0 \),
\[
0 \leq f(0) \leq f(s) + f'(s)(0 - s) \leq f(s) - p_\infty s \leq A_0 b(\tau) + \ell(sz) - p_\infty s = A_0 b(\tau),
\]
by (3.18) and the definition \( f(s) = \psi(sz, \tau) \). Note that as super-differential, \( z \cdot \partial \psi(z, \tau) \subset f'(1) \). Hence, setting \( s = 1 \) we obtain the assertion of the Lemma. \( \square \)

**Theorem 2.** Suppose \( \kappa > 0 \). Then there exist a function \( u \) and a constant \( M \) such that
\[
|\partial_\tau \psi(z, \tau)| + |\psi(z, \tau) - u(z)| \leq M(1 + r\tau)e^{-r\tau} \quad \forall \tau > 0, z \in \mathbb{R}^n.
\]
In addition, \( u \) is a Lipschitz continuous concave viscosity solution of the following equation
\[
\min \{-A[u], B(\nabla u)\} = 0, \quad 0 \leq u - \ell \leq \frac{A_0}{r} \quad \text{in} \quad \mathbb{R}^n
\]
where \( \ell(z) \) and \( B \) are as defined in (2.4) and (3.13) respectively, and
\[
A[\phi] := \frac{1}{2} \sum_{i,j} \sigma_{ij} z_i z_j (\partial_{z_i z_j} \phi - \partial_{z_i} \phi \partial_{z_j} \phi) + \sum_i \alpha_i z_i \partial_{z_i} \phi - r \phi.
\]

The estimate (4.3) implies that the finite horizon value function \( \psi \) and its time-derivative \( \partial_\tau \psi \) converge at an exponential rate governed by the interest rate \( r > 0 \), instead of the discount factor \( \beta \). If \( r \) goes to 0, the analytic optimal strategy in the absence of transaction costs indicates that the dollar values invested in risky assets and consumption tend to infinity [see, e.g., Merton (1969) and Chen et al. (2012)]. 5 It is also worthwhile pointing out that the solution \( u \) to the equation (4.4) can be formally regarded as the value function associated with the infinite horizon utility maximization problem. 6

**Proof of Theorem 2.** Keeping in mind that there are comparison principles for viscosity solutions, in the sequel, we treat the viscosity solution \( \psi \) as if it were a classical solution.

---

5 It should be admitted that the model with the CARA utility discussed here is less plausible than the one with the CRRA utility.

6 As mentioned in the introduction part, we first consider the finite horizon problem because the terminal utility of bequest can prevent unlimited consumption. If we directly consider the infinite horizon problem, some technical conditions should be given [see, e.g., Liu (2004)].
For any constant $h > 0$ and smooth function $W(\cdot)$, consider the function $\psi(z, \tau) := \psi(z, \tau + h) + W(\tau)$. One can calculate

$$
\psi_r(z, \tau + h) - A^T \psi(z, \tau + h) = \hat{\psi}_r - A^T \hat{\psi} - W'(\tau) - \kappa \xi(\tau + h) W(\tau)
$$

$$
= \hat{\psi}_r - A^T \hat{\psi} + [\kappa \xi(\tau) - \kappa \xi(\tau + h)](z \cdot \partial \hat{\psi} - \hat{\psi}) - W'(\tau) - \kappa \xi(\tau + h) W(\tau)
$$

$$
= \psi_r - A^T \hat{\psi} + f^W,
$$

where

$$
f^W(z, \tau) := -[W'(\tau) + \kappa \xi(\tau) W(\tau)] - A^T \hat{\psi} + f^W, B(\nabla \hat{\psi}(z, \tau)) = 0. \quad (4.6)
$$

(i) Suppose $0 < r \leq \kappa$. Then $\xi'(\tau) = (r - \kappa) \xi^2 e^{-rt} \leq 0$.

(1) Setting $W = 0$ and using the first inequality of (4.2) we find that

$$
f^W(z, \tau) = \kappa [\xi(\tau) - \xi(\tau + h)](z \cdot \partial \hat{\psi}(z, \tau + h) - \hat{\psi}(z, \tau + h)) \leq 0.
$$

This implies from (4.6) that $\hat{\psi}(z, \tau) := \psi(z, \tau + h)$ satisfies

$$
\min \{ \psi_r - A^T \hat{\psi}, B(\nabla \hat{\psi}(z, \tau)) \} \geq 0.
$$

Also $\hat{\psi}(z, 0) = \psi(z, h) \geq \ell(z) = \psi(z, 0)$. Hence, $\hat{\psi}(z, \tau) := \psi(z, \tau + h)$ is a supersolution, so by comparison principle we have $\psi(z, \tau) \leq \hat{\psi}(z, \tau + h)$.

(2) Setting $W = -A_0[b(\tau + h) - b(\tau)]$ and using $b' + \kappa \xi b = 1$ and the second inequality in (4.2), we obtain,

$$
f^W(z, \tau) = \kappa [\xi(\tau) - \xi(\tau + h)](z \cdot \partial \hat{\psi}(z, \tau + h) - \hat{\psi}(z, \tau + h) + A_0 b(\tau + h)) \geq 0.
$$

Thus, $\hat{\psi}(z, \tau) := \psi(z, \tau + h) - A_0[b(\tau + h) - b(\tau)]$ satisfies

$$
\min \{ \psi_r - A^T \hat{\psi}, B(\nabla \hat{\psi}(z, \tau)) \} \leq 0.
$$

Also, by (3.18), $\hat{\psi}(z, 0) = \psi(z, h) - A_0 b(h) \leq \ell(z)$. Hence, $\hat{\psi}$ is a subsolution, so $\psi(z, \tau) \geq \psi(z, \tau + h) - A_0[b(\tau + h) - b(\tau)]$.

In conclusion, for every $h > 0$,

$$
0 \leq \psi(z, \tau + h) - \psi(z, \tau) \leq A_0[b(\tau + h) - b(\tau)]. \quad (4.7)
$$

Sending $h \searrow 0$ we obtain

$$
0 \leq \psi_r(z, \tau) \leq A_0 b'(\tau) = O(1)(1 + \tau) e^{-rt}. \quad (4.8)
$$

Thus, $u(z) := \lim_{r \to \infty} \psi(z, \tau)$ exists, and by (3.18), $0 \leq u - \ell \leq A_0 b(\infty) = A_0/r$. Finally, sending $h \to \infty$ we obtain from (4.7) that

$$
0 \leq u(z) - \psi(z, \tau) \leq A_0[b(\infty) - b(\tau)] = O(1)(1 + \tau) e^{-rt}.
$$

This proves (4.3). Sending $\tau \to \infty$ in (3.19) and using (4.8), we find that $u$ is a Lipschitz continuous and concave viscosity solution of (4.4).

(ii) Suppose $r \geq \kappa$. Then $\xi'(\tau) = (r - \kappa) \xi^2 e^{-rt} \geq 0.$
find that exists. Also, sending section remain valid in the higher dimensional case. In particular, this implies that \( \psi(z, h) \leq \psi(z, \tau) + \psi(z, \tau + h) \in A_0b(h)Z(\tau) \). Also, \( \hat{\psi}(z, 0) = \psi(z, h) - A_0b(h) \leq \phi_0(z) \). Hence, \( \hat{\psi} = \psi(z, \tau) - A_0b(h)Z(\tau) \) is a subsolution, so \( \psi(z, \tau + h) - A_0b(h)Z(\tau) \leq \psi(z, \tau) \).

In summary, we have

\[
A_0[b(\tau + h) - b(\tau) - b(h)Z(\tau)] \leq \psi(z, \tau + h) - \psi(z, \tau) \leq A_0b(h)Z(\tau).
\]

Sending \( h \downarrow 0 \) we obtain

\[
A_0[b'(\tau) - Z(\tau)] \leq \psi(z, \tau) \leq A_0Z(\tau).
\]

In particular, this implies that \( |\psi| = O(1)(1 + \tau)e^{-\tau} \). Consequently, \( u = \lim_{\tau \to \infty} \psi(z, \tau) \) exists. Also, sending \( h \to \infty \) we obtain from (4.9) that

\[
A_0 \left[ \frac{1}{r} - b(\tau) - \frac{1}{r}Z(\tau) \right] \leq u(z) - \psi(z, \tau) \leq \frac{A_0}{r}Z(\tau).
\]

This implies that \( |\psi(z, \tau) - u(z)| = O(1)(1 + \tau)e^{-\tau} \). Finally, sending \( \tau \to \infty \) in (3.19), we find that \( u \) is a viscosity solution of (4.4). This completes the proof of Theorem 2.

5. Shape and Location of The Trading/No-Trading Zones

In this section, we investigate the shape and location of the trading/no-trading regions in the two dimensional case for the infinite horizon problem.\(^7\) The rigorous analysis in this section needs \( C^1 \) regularity of \( u \) whose proof is very technical, and therefore is given in the next section. Regarding the shape, we have the following result.

**Theorem 3.** Let \((i, i) = (1, 2)\) or \((2, 1)\). Define

\[
B_i := \{ z | \partial_z u(z) = 1 + \lambda_i \},
\]

\[
S_i := \{ z | \partial_z u(z) = 1 - \mu_i \},
\]

\[
N_i := \{ z | 1 - \mu_i < \partial_z u(z) < 1 + \lambda_i \},
\]

and denote \( SS = S_1 \cap S_2, SN = S_1 \cap N_2, SB = S_1 \cap B_2, NS = N_1 \cap S_2, NT = N_1 \cap N_2, NB = N_1 \cap B_2, BS = B_1 \cap S_2, BN = B_1 \cap N_2, \) and \( BB = B_1 \cap B_2 \). Then

(1) There are bounded functions \( l_i^\pm(\cdot) \) such that

\[
B_i = \{ (z_1, z_2) | z_1 \in \mathbb{R}, \ z_i \leq l_i^-(z_i) \},
\]

\[
S_i = \{ (z_1, z_2) | z_1 \in \mathbb{R}, \ z_i \geq l_i^+(z_i) \},
\]

\[
N_i = \{ (z_1, z_2) | l_i^- < z_i < l_i^+(z_i) \}.
\]

(2) Each intersection \( \partial S_1 \cap \partial S_2, \partial S_1 \cap \partial N_2, \partial B_1 \cap \partial S_2, \) and \( \partial B_1 \cap \partial B_2 \) is a singleton, so the four boundaries \( \partial S_1, \partial B_1, \partial S_2, \) and \( \partial B_2 \) divide the plane into nine regions, with open region \( NT \) in the center surrounded in clockwise order by closed regions \( SS, SN, SB, NB, BS, BN, \) and \( NS \).

\(^7\)We consider the two dimensional case because of notational simplicity. We believe that all results of this section remain valid in the higher dimensional case.
(3) The boundary of each of corner regions SS, SB, BB, and BS consists of one vertical and one horizontal half line, whereas the boundary of each of SN, NS, BN, and NB consists of two parallel either vertical or horizontal half lines and a curve in between connecting the end points of the two half lines; c.f. Figure 1.

The theorem implies the following: There are intervals \([b_i^\pm, s_i^\pm]\) such that

\[
\begin{align*}
SS &= ([s_1^+, \infty) \times [s_2^+, \infty), \\
SN &= \{(z_1, z_2) | z_2 \in (b_2^+, s_2^+), z_1 \geq l_2^+(z_2)\}, \\
SB &= ([s_1^-, \infty) \times (-\infty, b_2^+], \\
NB &= \{(z_1, z_2) | z_1 \in (b_1^-, s_1^+), z_2 \leq l_1^-(z_1)\}, \\
BB &= (-\infty, b_1^-) \times (-\infty, b_2^-], \\
BN &= \{(z_1, z_2) | z_2 \in (b_2^-, s_2^-), z_1 \leq l_2^-(z_2)\}, \\
BS &= (-\infty, b_1^+] \times [s_2^-, \infty), \\
NS &= \{(z_1, z_2) | z_1 \in (b_1^+, s_1^+), z_2 \geq l_1^+(z_1)\}.
\end{align*}
\]

The no-trading region \(NT\) is bounded by four curves: \(\Gamma_{2+}\) from the right, \(\Gamma_{2-}\) from the left, \(\Gamma_{1+}\) from the top, and \(\Gamma_{1-}\) from the bottom where

\[
\Gamma_{2\pm} := \{(l_2^+(z_2), z_2) | z_2 \in (b_2^+, s_2^+)\}, \quad \Gamma_{1\pm} := \{(l_1^+(z_1), z_1) \in (b_1^+, s_1^+)\}.
\]

These four curves connect each other only at their tips:

\[
\begin{align*}
l_1^+(b_1^+) &= \lim_{z_i \downarrow b_1^+} l_1^+(z_i), & l_1^+(s_1^+) &= \lim_{z_i \downarrow s_1^+} l_1^+(z_i), \\
(b_1^+, l_1^+(b_1^+)) &= (l_2^-(s_2^-), s_2^-), & (l_2^+(s_2^+), s_2^+) &= (s_1^+, l_1^+(s_1^-)), \\
(l_2^-(b_2^-), b_2^-) &= (b_1^+, l_1^+(b_1^+)), & (s_1^-, l_1^+(s_1^-)) &= (l_2^+(b_2^-), b_2^-).
\end{align*}
\]

Each function \(l_i^\pm\) is constant outside \((b_i^\pm, s_i^\pm)\):

\[
l_i^+(z_i) = l_i^+(s_i^+) \quad \forall z_i \leq b_i^+, \quad l_i^+(z_i) = l_i^+(s_i^+) \quad \forall z_i \geq s_i^+.
\]

As emphasized in the introduction part, Theorem 3 is very much needed for the trading strategy to be well-defined. It is well-known that except at the initial time, transactions occur at the boundary of the no-trading region. When the initial portfolio falls out of the no-trading region, the shape of the trading regions stated in the theorem implies a unique trading strategy to move the portfolio to the boundary of the no-trading region. However, at this stage we cannot prove the smoothness of the curves \(l_i^\pm\), so we are unable to construct the controlled portfolio process as a reflected diffusion as in Davis and Norman (1990) and Shreve and Soner (1994).

Regarding the location, we have the following

**Theorem 4.** Let \((m_1, m_2)\) and \(A_0\) be as given in (3.14) with \(n = 2\), and

\[
\begin{align*}
\sigma_1 &= \sqrt{\sigma_{11}}, \quad \sigma_2 = \sqrt{\sigma_{22}}, \quad \rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}, \\
M(h) &= \frac{\sqrt{2(A_0 - rh)}}{1 - \rho^2}.
\end{align*}
\]
Also, define the ellipse \( C(k_1, k_2, h) \) and constant \( c_{ij} \) by
\[
C(k_1, k_2, h) := \left\{ (z_1, z_2) \left| \frac{1}{2} \sum_{i,j} \sigma_{ij}(k_iz_i - m_i)(k_jz_j - m_j) = A_0 - rh \right. \right\},
\]
(5.4)
\[
c_{ij} = \lim_{(z_1, z_2) \to (-1)^i z_1 \to -\infty, (-1)^j z_2 \to -\infty} \{ u(z) - \ell(z) \}, \quad i, j = 1, 2.
\]
(5.5)

1. The no-trading region \( \text{NT} \) is contained in the set
\[
D := \bigcup_{1 - \mu_i < k_1 < 1 + \lambda_i, h \geq u(0)} C(k_1, k_2, h).
\]
(5.6)

2. The corner \((s_1^+, s_2^+)\) of \( SS \) lies on the ellipse \( C(1 - \mu_1, 1 - \mu_2, c_{22}) \) and on its top-left part in the sense that
\[
-\rho \leq \frac{(1 - \mu_1)s_1^+ - m_1}{M(c_{22})/\sigma_1} \leq 1, \quad -\rho \leq \frac{(1 - \mu_2)s_2^+ - m_2}{M(c_{22})/\sigma_2} \leq 1.
\]

3. The corner \((s_1^-, b_2^+)\) of \( SB \) lies on the bottom-right part of \( C(1 - \mu_1, 1 + \lambda_2, c_{21}) \):
\[
\rho \leq \frac{(1 - \mu_1)s_1^- - m_1}{M(c_{21})/\sigma_1} \leq 1, \quad -1 \leq \frac{(1 + \lambda_2)b_2^+ - m_2}{M(c_{21})/\sigma_2} \leq -\rho.
\]

4. The corner \((b_1^-, b_2^+)\) of \( BB \) lies on bottom-left part of \( C(1 + \lambda_1, 1 + \lambda_2, c_{11}) \):
\[
-1 \leq \frac{(1 + \lambda_1)b_1^- - m_1}{M(c_{11})/\sigma_1} \leq \rho, \quad -1 \leq \frac{(1 + \lambda_2)b_2^- - m_2}{M(c_{11})/\sigma_2} \leq \rho.
\]

5. The corner \((b_1^+, s_2^-)\) of \( BS \) lies on top-left of \( C(1 + \lambda_1, 1 - \mu_2, c_{12}) \):
\[
-1 \leq \frac{(1 + \lambda_1)b_1^+ - m_1}{M(c_{12})/\sigma_1} \leq -\rho, \quad \rho \leq \frac{(1 - \mu_2)s_2^- - m_2}{M(c_{12})/\sigma_2} \leq 1,
\]

We remark that \( 0 \leq u(0) \leq c_{ij} \leq \frac{A_0}{\rho} \) since \( 0 \leq u(z) - \ell(z) \leq A_0/\rho \) and \( u(z) - \ell(z) \) is an increasing concave function in each radial direction.

The theorem shows that the no-trading region is contained in a union of uniformly bounded ellipses. Moreover, the location of the corners of the no-trading region is estimated. Hence we can restrict attention to a bounded domain to study the problem. In particular, this allows us to do computations in a bounded domain.

5.1. The Bound of the No-Trading Region.

For each \( z^0 \), we define the first order approximation of \( u \) by
\[
\pi(z^0, z) := u(z^0) + \nabla u(z^0) \cdot (z - z^0).
\]
By concavity, we have
\[
u(z) \leq \pi(z^0, z),
\]
which will be often used later on. In particular,
\[
u(0) \leq \pi(z^0, 0) \leq u(z^0) - \nabla u(z^0)z^0.
\]
(5.7)
We now decompose $A[u]$ as $A[u] = \mathcal{L}u - f$ where

$$
\mathcal{L}u(z) = \frac{1}{2} \sum_{i,j} \sigma_{ij} z_i z_j \partial_{z_i z_j} u(z),
$$

(5.8)

and

$$
f(z) = \frac{1}{2} \sum_{i,j} \sigma_{ij} z_i z_j \partial_{z_i} u(z) \partial_{z_j} u(z) - \sum_i \alpha_i z_i \partial_{z_i} u(z) + ru(z)
$$

$$
= \frac{1}{2} \sum_{i,j} \sigma_{ij} (z_i \partial_{z_i} u - m_i)(z_j \partial_{z_j} u - m_j) - A_0 + r(u - z \cdot \nabla u).
$$

(5.9)

Thanks to Theorem 5 (in the next section), $f$ is continuous. If $z \in \mathcal{N}$, then $A[u](z) = 0$ so $\mathcal{L}u(z) = f(z)$. Using concavity of $u$ we derive

$$
0 \geq \mathcal{L}u(z) = f(z) \geq \frac{1}{2} \sum_{i,j} \sigma_{ij} (z_i \partial_{z_i} u - m_i)(z_j \partial_{z_j} u - m_j) - A_0 + ru(0),
$$

where we have used (5.7) in the last inequality. Thus, $z$ is on or inside of the ellipse $C(\partial_{z_1} u, \partial_{z_2} u, u(0))$. Since $\partial_{z_i} u(z) \in (1 - \mu_i, 1 + \lambda_i)$ we see that $z \in \mathcal{D}$ defined in (5.6). The first assertion the of Theorem 4 thus follows. \qed

In the sequel, we shall use the following important fact:

**Lemma 5.1.** If $z^0 = (z^0_1, z^0_2) \in S_1$ then $[z^0_1, \infty) \times \{z^0_2\} \in S_1$ and $u(z) = \pi(z^0, z)$ and $\nabla u(z) = \nabla u(z^0)$ for all $z \in [z^0_1, \infty) \times \{z^0_2\}$. Analogous assertions hold also for the cases $z^0 \in B_1$, $z^0 \in S_2$, and $z^0 \in B_2$, respectively.

**Proof.** Suppose $z^0 = (z^0_1, z^0_2) \in S_1$, i.e., $\partial_{z_1} u(z^0) = 1 - \mu_1$. Since $\partial_{z_1} z_i u \leq 0$ and $\partial_{z_i} u \geq 1 - \mu_1$, we have $\partial_{z_1} u(z_1, z^0_2) = 1 - \mu_1$ for all $z_1 \geq z^0_1$; that is, $[z^0_1, \infty) \times \{z^0_2\} \in S_1$. In addition, for each $z = (z_1, z^0_2)$ with $z_1 \geq z^0_1$, $u(z) = u(z_0) + (1 - \mu_1)(z_1 - z^0_1) = \pi(z_0, z)$. Since $u(z) \leq \pi(z_0, z)$ for all $z \in \mathbb{R}^2$, we also have $\nabla u(z) = \nabla \pi(z^0, z) = \nabla u(z^0)$ for every $z \in [z^0_1, \infty) \times \{z^0_2\}$.

The other cases can be similarly proven. This completes the proof. \qed

5.2. Limit Profile.

With $k^+_i := 1 - \mu_i$ and $k^-_i := 1 + \lambda_i$, we define the limit profile by

$$
v^+_2(z_2) := \lim_{z_1 \to \pm \infty} (u(z_1, z_2) - k^+_1 z_1), \quad v^+_1(z_1) := \lim_{z_2 \to \pm \infty} (u(z_1, z_2) - k^+_2 z_2).
$$

(5.10)

**Lemma 5.2.**

1. With $k^+_i := 1 - \mu_i$ and $k^-_i := 1 + \lambda_i$, $v^+_i$ defined in (5.10) is concave and

$$
u(z_1, z_2) \leq \min \left\{ v^+_2(z_2) + k^+_1 z_1, v^+_1(z_1) + k^+_2 z_2 \right\} \quad \forall z \in \mathbb{R}^2.
$$

2. There are intervals $(b^+_i, s^+_i)$ and functions $l^+_i$ defined on $(b^+_i, s^+_i)$ such that

$$
v^+_i(z_i) \begin{cases} 
= 1 - \mu_i & \text{if } z_i \geq s^+_i, \\
\in (1 - \mu_i, 1 + \lambda_i) & \text{if } z_i \in (b^+_i, s^+_i), \\
= 1 + \lambda_i & \text{if } z_i \leq b^+_i,
\end{cases}
$$
\[ u(z_1, z_2) = \begin{cases} 
\quad v^+_i(z_2) + (1 - \mu_1)z_1 & \text{if } z_1 \geq l^+_i(z_2), z_2 \in (b_i^+, s_i^+), \\
\quad v^-_i(z_2) + (1 + \lambda_1)z_1 & \text{if } z_1 \leq l^-_i(z_2), z_2 \in (b_i^-, s_i^-), \\
\quad v^+_1(z_1) + (1 - \mu_2)z_2 & \text{if } z_2 \geq l^+_1(z_1), z_1 \in (b_1^+, s_1^+), \\
\quad v^-_1(z_1) + (1 + \lambda_2)z_2 & \text{if } z_2 \leq l^-_1(z_1), z_1 \in (b_1^-, s_1^-), 
\end{cases} \]

\[
\partial_z u(z_1, z_2) > 1 - \mu_1 \quad \text{if } z_1 < l^+_i(z_2), z_2 \in (b^+_i, s^+_i),
\]

\[
\partial_z u(z_1, z_2) < 1 + \lambda_1 \quad \text{if } z_1 > l^-_i(z_2), z_2 \in (b^-_i, s^-_i),
\]

\[
\partial_z u(z_1, z_2) > 1 - \mu_2 \quad \text{if } z_2 < l^+_1(z_1), z_1 \in (b^+_1, s^+_1),
\]

\[
\partial_z u(z_1, z_2) < 1 + \lambda_2 \quad \text{if } z_2 > l^-_1(z_1), z_1 \in (b^-_1, s^-_1).
\]

(3) Define
\[
l^+_i(s^+_i) = \lim_{z_i \rightarrow \infty} l^+_i(z_i), \quad l^-_i(s^-_i) = \lim_{z_i \rightarrow -\infty} l^-_i(z_i),
\]

\[
l^+_1(b^+_1) = \lim_{z_1 \rightarrow b_1^+} l^+_1(z_1), \quad l^-_1(b^-_1) = \lim_{z_1 \rightarrow -b_1^-} l^-_1(z_1).
\]

Then
\[
[l^+_2(s^+_2), l^-_2(s^-_2)] \cup \{s^+_2\} \cup \{s^-_2\} \subset \partial NT \cap SS,
\]

\[
[l^+_1(s^+_1), l^-_1(s^-_1)] \subset \partial NT \cap SB,
\]

\[
[l^+_2(s^+_2), l^-_2(s^-_2)] \cup \{s^+_2\} \cup \{s^-_2\} \subset \partial NT \cap BS,
\]

\[
[l^+_2(b^+_2), l^-_2(b^-_2)] \cup \{b^+_2\} \cup \{b^-_2\} \subset \partial NT \cap BB.
\]

**Proof.** By symmetry, we need only consider the function \( v := v^+_2 \).

(1) Note that \( u(z_1, z_2) - (1 - \mu_1)z_1 \) is a concave function, \( \partial_{z_1}[u(z_1, z_2) - (1 - \mu_1)z_1] = \partial_{z_1}u - (1 - \mu_1) \geq 0 \), and \( 0 \leq u(z) - l(z) \leq A_0/r \). Hence, \( v(z_2) := \lim_{z_1 \rightarrow -\infty}[u(z_1, z_2) - (1 - \mu_1)z_1] \) exists, \( v \) is concave, and \( v(z_2) \geq u(z_1, z_2) - (1 - \mu_1)z_1 \).

(2) Let \( z_2 \in \mathbb{R} \) be a generic point such that \( 1 - \mu_2 < v'(z_2) < 1 + \lambda_2 \). Since \( \partial_{z_2}u(z_1, z_2) \rightarrow v'(z_2) \) as \( z_1 \rightarrow \infty \), \( \partial_{z_2}u(z_1, z_2) \) is \( (1 - \mu_2, 1 + \lambda_2) \) for all \( z_2 \gg 1 \). As \( NT \) is bounded, we must have \( \partial_{z_2}u(z_1, z_2) = 1 - \mu_1 \) for all \( z_1 \gg 1 \). In addition, since \( \partial_{z_1}u \rightarrow 1 + \lambda_1 \) as \( z_1 \rightarrow -\infty \), we can define
\[
l(z_2) := \min\{z_1 \in \mathbb{R} \mid \partial_{z_1}u(z_1, z_2) = 1 - \mu_1\}.
\]

Since \( u(\cdot, z_2) \) is concave, we must have
\[
\partial_{z_1}u(z_1, z_2) < 1 - \mu_1 \quad \forall z_1 < l(z_2), \quad \partial_{z_1}u(z_1, z_2) = 1 - \mu_1 \quad \forall z_1 \geq l(z_2).
\]

Denote \( z^0 := (l(z_2), z_2) \). Then by Lemma 5.1, \( u(z) = \pi(z^0, z) = v(z^0) + (1 - \mu_1)z_1 \) and \( \nabla u(z) = \nabla v(z^0) \) for each \( z \in [l(z_2), \infty) \times \{z_2\} \). Consequently, \( \partial_{z_2}u(z_1, z_2) = v'(z_2) \in (1 - \mu_2, 1 + \lambda_2) \) for all \( z_1 \in [l(z_2), \infty) \). Thus, for small positive \( \varepsilon \), \( B(\nabla v(z)) > 0 \) on \( [l(z_2) - \varepsilon, l(z_2)] \times \{z_2\} \). Hence, \( [l(z_2) - \varepsilon, l(z_2)] \times \{z_2\} \in NT \) and \( (l(z_2), z_2) \in \partial NT \). Since \( NT \) is bounded and \( v \) is concave, there exist bounded \( b \) and \( s \) such that \( v' = 1 + \lambda_2 \) on \( (-\infty, b], \quad 1 + \lambda_2 > v' > 1 - \mu_2 \) on \( (b, s), \quad v' = 1 - \mu_2 \) on \( [s, \infty) \).

(3) Now we define \( l(s) = \lim_{z_2 \rightarrow s} l(z_2), \ l_+(s) = \lim_{z_2 \rightarrow s^+} l(z_2), \) and \( z^* = (l(s), s) \). By continuity, we have \( \partial_{z_1}u(l(s), s) = 1 - \mu_1 \). This implies that \( \nabla u(z) = (1 - \mu_1, v'(s)) = (1 - \mu_1, 1 - \mu_2) \) for all \( z \in [l(s), \infty) \times \{s\} \). Hence, \( [l(s), \infty) \times \{s\} \in SS \).
Note that for each \( z_2 < s \) and \( z_1 \in [l(s), \infty) \), we have
\[
    u(z_1, s) - u(z_1, z_2) = v(s) - v(z_2) - \int_{z_2}^{s} [\partial_{z_1} u(\xi, s) - \partial_{z_2} u(\xi, z_2)] d\xi
    = \int_{z_2}^{s} v'(y) dy + \int_{z_2}^{\infty} [\partial_{z_1} u(\xi, z_2) - (1 - \mu_1)] d\xi > (1 - \mu_2)(s - z_2).
\]
It follows by concavity that \( \partial_{z_1} u(z_1, z_2) > 1 - \mu_2 \). Recalling \( (s, l) = (s_2^+, l_2^+) \), we have
\[
    \partial_{z_1} u > 1 - \mu_2 \quad \text{on} \quad [l_2^+(s), \infty) \times (-\infty, s_2^+). \tag{5.11}
\]
It then follows by the definition of \( l(s) \) and \( l^*(s) \) that \([l(s), l^*(s)] \times \{ s \} \subset \partial NT\).

Similarly we can work on the other functions \( v_i' \) to complete the proof of Lemma 5.2. \( \square \)

5.3. The Intersection of \( \partial NT \) with \( B_1 \cap B_2, S_1 \cap S_2 \) and \( B_i \cap S_j \).

In this subsection we prove the following:

Lemma 5.3. Let \( c_{ij} \) and \( C(k_1, k_2, h) \) be defined as in (5.5) and (5.4).

1. The set \( \partial NT \cap SS \) is a single point on top-right of the ellipse \( C(1 - \mu_1, 1 - \mu_2, c_{22}) \).
2. The set \( \partial NT \cap SB \) is a single point on bottom-right of \( C(1 - \mu_1, 1 + \lambda_2, c_{22}) \). If \( (z_1^0, z_2^0) \in SB \), then \([z_1^0, \infty) \times (-\infty, z_2^0] \subset SB \).
3. The set \( \partial NT \cap BB \) is a single point on bottom-left of \( C(1 + \lambda_1, 1 + \lambda_2, c_{22}) \). If \( (z_1^0, z_2^0) \in BB \), then \((-\infty, z_1^0] \times (-\infty, z_2^0] \subset BB \).
4. The set \( \partial NT \cap BS \) is a single point on top-left of \( C(1 + \lambda_1, 1 - \mu_2, c_{12}) \). If \( (z_1^0, z_2^0) \in BS \), then \((-\infty, z_1^0] \times [z_2^0, \infty) \subset BS \).

Proof. (i) Suppose \( z^0 = (z_1^0, z_2^0) \in S_1 \cap S_2 \).

First we show that \([z_1^0, \infty) \times [z_2^0, \infty) \subset SS \). Indeed by Lemma 5.1, \( u(z) = \pi(z_0, z) \) on \(([z_1^0, \infty) \times [z_2^0, \infty)) \cup ([z_1^0] \times [z_2^0, \infty)) \). This implies that \( u(\cdot) \geq \pi(z_0, \cdot) \) on \([z_1^0, \infty) \times [z_2^0, \infty) \) since \( \pi(z_0, \cdot) \) is linear and \( u(\cdot) \) is concave. On the other hand, we have \( u(z) \leq \pi(z_0, z) \) on \( \mathbb{R}^2 \). Thus we must have \( u(z) = \pi(z_0, z) \) and \( \nabla u(z) = \nabla u(z_0) = (1 - \mu_1, 1 - \mu_2) \) so \([z_1^0, \infty) \times [z_2^0, \infty) \subset SS \). In addition, on \([z_1^0, \infty) \times [z_2^0, \infty) \),
\[
    u(z) = u(z_0) + (1 - \mu_1)(z_1 - z_0^1) + (1 - \mu_2)(z_2 - z_0^2) = c_{22} + z \cdot \nabla u(z_0).
\]

Next since \( -A[u] = f(z) - Lu \geq 0 \) on \( \mathbb{R}^2 \), using the linearity of \( u \) on \([z_1^0, \infty) \times [z_2^0, \infty) \) we obtain
\[
    0 = \lim_{s \to 0} Lu(z^0 + se_1 + se_2) \leq \lim_{s \to 0} f(z^0 + se_1 + se_2) = f(z^0).
\]
Using \( u = c_{22} + z \cdot \nabla u(z_0) \) on \([z_1^0, \infty) \times [z_2^0, \infty) \) we find that \( f(z^0) = f_{22}(z^0) \geq 0 \) where
\[
    f_{22}(z) := \frac{1}{2} \sum_{i,j} \sigma_{ij} (z_1[1 - \mu_1] - m_i)(z_1[1 - \mu_1] - m_j) - A_0 + rc_{22} \quad \forall z \in \mathbb{R}^2.
\]
This analysis in particular implies that \( f_{22}(z) \geq 0 \) for every \( z \in SS \).

(ii) Next, suppose \( z^0 \in \partial NT \cap SS \). Note that when \( z \in NT \), we have \( 0 = -A[u](z) = f(z) - Lu[z] \), i.e. \( f(z) = Lu[z] \leq 0 \) (as \( u \) is concave). Hence,
\[
    f(z^0) = \lim_{z \to z^0} f(z) = \lim_{z \to z^0} Lu[z] \leq 0.
\]
Thus, we must have $f(z^0) = 0$, i.e., $z^0 \in C := C(1-\mu_1,1-\mu_2,c_{22})$. Moreover, since $f_{22}(z) \geq 0$ for every $z \in [z_1^0, \infty) \times [z_2^0, \infty)$ and $f(z) < 0$ for each $z$ inside the ellipse $C$, we see that $z^0$ lies on the top-right part of the ellipse $C$. Locating the highest and rightmost points of the ellipse $C$, we then derive that

$$f_{22}(z^0) = 0, \quad -\rho \leq \frac{(1-\mu_1)z_1^0 - m_1}{M(c_{22})/\sigma_1} \leq 1, \quad -\rho \leq \frac{(1-\mu_2)z_2^0 - m_2}{M(c_{22})/\sigma_2} \leq 1.$$  

(iii) Now we show that $\partial NT \cap SS$ is a singleton, by a contradiction argument. Suppose $z^0 = (z_1^0, z_2^0) \in \partial NT \cap SS$. $z^0 = (\hat{z}_1^0, \hat{z}_2^0) \in \partial NT \cap SS$ and $\hat{z}_1^0 \neq z^0$. Then both $z^0$ and $\hat{z}_1^0$ lies on the upper-right part of the ellipse $C(1-\mu_1,1-\mu_2,c_{22})$. Exchanging the roles of $z^0$ and $\hat{z}_1^0$, we can assume that $z_2^0 < \hat{z}_2^0$ and $z_1^0 > \hat{z}_1^0$. Note that $u(z) = \pi(z^0, z)$ for all $z \in [\hat{z}_1^0, \infty) \times [\hat{z}_2^0, \infty)$ and $u(z) = \pi(z, z)$ for all $z \in [z_1^0, \infty) \cap [z_2^0, \infty)$. Hence, $\pi(z^0, z) = \pi(z, z)$ for all $z \in \mathbb{R}^2$. On the other-hand, if $u(z) = \pi(z^0, z)$, then $\nabla u(z) = \nabla \pi(z^0, z) = \nabla u(z_0)$ so $z \in SS$. Therefore,

$$SS := \{ z \in \mathbb{R}^2 \mid \nabla u(z) = (1-\mu_1,1-\mu_1) \} = \{ z \in \mathbb{R}^2 \mid u(z) = \pi(z^0, z) \}.$$  

Note that $u$ is concave, $\pi(z^0, \cdot)$ is linear, and $u(\cdot) \leq \pi(z^0, \cdot)$ on $\mathbb{R}^n$. We derive that $SS$ is a convex set. Consequently, it contains $L$, the line segment connecting $z^0$ and $\hat{z}_1^0$.

Next for each $s \in \mathbb{R}$, denote $z^s = (z_1^s, z_2^s) + (s, s)$. For $s < 0$, $z^s \notin SS$ since otherwise it would imply $[\hat{z}_1^0 + s, \infty) \times [\hat{z}_2^0 + s, \infty) \in SS$, contradicting $z^0 \in \partial NT$. Hence, there exists $s^* \geq 0$ such that $z^s = (\hat{z}_1^0 + s^*, \hat{z}_2^0 + s^*) \in SS$. The point $z^s$ lies on or below $L$ so is an interior point of the ellipse $C$. There are two cases: (a) $s^* > 0$, and (b) $s^* = 0$.

Consider case (a) $s^* > 0$. Then for each $s \in [0, s^*)$, $z^s \notin SS$ since $SS$ is convex. Also $\partial_{z^1} u(z^s) > 1 - \mu_1$ since by Lemma 5.1, $\partial_{z^1} u(z^s) = 1 - \mu_1$ would imply that $\nabla u(z^s) = \nabla u(z_0)$ where $z$ is the intersection of $L$ with the line $z_2 = z_2^0 + s$. Similarly, $\partial_{z^2} u(z^s) > 1 - \mu_2$, for each $s \in [0, s^*)$. Thus, $z^s \in NT$ for all $s \in [0, s^*)$. This means that $z^s \in \partial NT \cap SS$, which is impossible since $z^s \notin C$.

Consider case (b) $s^* = 0$. First of all $\partial_{z^1} u(z_1^0, z_2^0) > 1 - \mu_1$ for all $z_1 < z_1^0$ since otherwise it would imply $\nabla u(z_1^0, z_2^0) = \nabla u(z^0)$ and thus $[z_1, \infty) \times [z_2^0, \infty) \in SS$ contradicting $z^0 \in \partial NT$. Similarly, $\partial_{z^2} u(z_1^0, z_2^0) > 1 - \mu_2$ for every $z_2 < z_2^0$. Now for every small positive $\varepsilon$, consider the closed set

$$D_\varepsilon := [z_1^0 - \varepsilon, z_1^0] \times [z_2^0 - \varepsilon, z_2^0] \setminus (z_1^0 - \varepsilon/2, z_1^0) \times (z_2^0 - \varepsilon/2, z_2^0).$$  

If $z^* = (z_1^*, z_2^*) \in SS \cap D_\varepsilon$ we would have $z_1^* < z_1^0$ and $z_2^* < z_2^0$ so $z^0$ is an interior point of $[z_1^*, \infty) \times [z_2^*, \infty) \subset SS$, a contradiction. Thus, $SS \cap D_\varepsilon = \emptyset$. Consequently, the closed sets $D_\varepsilon \cap S_1$ and $D_\varepsilon \cap S_2$ are disjoint. Since $D_\varepsilon$ cannot be written as the union of two disjoint closed proper subsets, the set $D_\varepsilon \setminus (S_1 \cup S_2)$ is non-empty. This means that $D_\varepsilon \cap NT \neq \emptyset$. Consequently, $(z_1^0, z_2^0) \in \partial NT \cap SS$, but this is impossible since $(\hat{z}_1^0, \hat{z}_2^0)$ does not lie on $C$. Therefore $\partial NT \cap SS$ is a singleton.

The proof for the singleness of $\partial NT \cap BS$, $\partial NT \cap SB$, and $\partial NT \cap BB$ is similar. This completes the proof of the Lemma.

5.4. Completion of the Proof of Theorems 3, 4.

By Lemmas 5.2 and 5.3, we see that the limits in (5.2) exist, and the limits satisfy the matching condition stated in (5.2). We extend $I_1^+ \,(b_i^+, s_i^\pm)$ to $\mathbb{R}$ by (5.3).
By the definition of \( l_1^+ \) on \((b_1^+, s_1^+)\) we know that when \( z_1 \in (b_1^+, s_1^+) \), \( \partial z_2 u(z_1, z_2) > 1 - \mu_2 \) if and only if \( z_1 < l_1^+(z_1) \). Also, in view of (5.11) and the matching (5.2), we derive that when \( z_1 \in [s_1^+, \infty) \), \( \partial z_2 u(z_1, z_2) > 1 - \mu_2 \) if and only if \( z_2 < l_1^+(s_1^+) = l_1^+(z_1) \). Similarly, we can show that when \( z_1 \in (-\infty, b_1^+] \), \( \partial z_2 u(z_1, z_2) < 1 - \mu_2 \) if and only if \( z_2 < l_1^+(b_1^+) = l_1^+(z_1) \). Thus,

\[
S_2 := \{ z \mid \partial z_2 u = 1 - \mu_2 \} = \{(z_1, z_2) \mid z_1 \in \mathbb{R}, z_2 \geq l_1^+(z_1) \}.
\]

Similarly, we can show the other equations in (5.1). The rest assertions of Theorems 3 and 4 thus follow from Lemmas 5.2 and 5.3.

6. \( C^1 \) Regularity

In this section we show that the viscosity solution \( u \) of the infinite horizon problem is \( C^1 \) except on the coordinates planes where the elliptic operator \( A \) is degenerate. The \( C^1 \) continuity plays a critical role in analyzing the optimal strategy, where a key step is to derive the continuity of \( f(\cdot) \) as defined in (5.9).

We begin with recalling the definition of a viscosity solution:

**Definition 1.** A function \( u \) defined on \( \mathbb{R}^n \) is called a viscosity solution of (4.4) if \( u \) is continuous, \( u - \ell \in L^\infty(\mathbb{R}^n) \), and the following holds:

1. If \( \zeta \) is a \( C^2 \) function in \( B_\varepsilon(z^0) := \{ z \in \mathbb{R}^n \mid |z - z^0| < \varepsilon \} \) for some \( z^0 \in \mathbb{R}^n \) and \( \varepsilon > 0 \), and that \( \zeta(z) - u(z) \geq 0 = \zeta(z^0) - u(z^0) \) for every \( z \in B_\varepsilon(z^0) \), then
   \[
   \min \left\{ -A[\zeta](z^0), \quad B(\nabla \zeta(z^0)) \right\} \leq 0.
   \]

2. If \( \zeta \) is a \( C^2 \) function in \( B_\varepsilon(z_0) := \{ z \in \mathbb{R}^n \mid |z - z_0| < \varepsilon \} \) for some \( z_0 \in \mathbb{R}^n \) and \( \varepsilon > 0 \), and that \( \zeta(z) - u(z) \leq 0 = \zeta(z_0) - u(z_0) \) for each \( z \in B_\varepsilon(z_0) \), then
   \[
   \min \left\{ -A[\zeta](z_0), \quad B(\nabla \zeta(z_0)) \right\} \geq 0.
   \]

One can show that viscosity solution of (4.4) is unique. Since the solution is the limit \( \psi(\cdot, \tau) \) that is concave, we see that viscosity solution of (4.4) is concave. We shall use this fact to prove the \( C^1 \) regularity of \( u \).

**Theorem 5.** Let \( u \) be the solution of (4.4). Then \( z_i \partial z_i u \in C(\mathbb{R}^n) \) for each \( i = 1, \cdots, n \); consequently, \( u \in C^1(\Omega) \) where \( \Omega = \{(z_1, \cdots, z_n) \in \mathbb{R}^n \mid \prod_i z_i \neq 0 \} \).

**Proof.** For illustration, we consider only the two space dimensional case.

First we show that \( \partial_1 u(z_1, z_2) \) is a singleton if \( z_1 \neq 0 \). We use a contradiction argument. Suppose the assertion is not true. Then there exist \( z^0 = (z_1^0, z_2^0) \) with \( z_1^0 \neq 0 \) and \( p = (p_1, p_2) \in \partial u(z^0) \) and \( q = (q_1, q_2) \in \partial u(z^0) \) with \( 1 - \mu_1 \leq p_1 < q_1 \leq 1 + \lambda_1 \).

By the definition of super-differential, we have

\[
\begin{aligned}
u(z) \leq \min \{ u(z^0) + p \cdot (z - z^0), \quad u(z^0) + p \cdot (z - z^0) \} \quad &\forall z \in \mathbb{R}^2. \\
\end{aligned}
\]

Now for any \( 0 < \varepsilon \ll 1 \), consider the quadratic concave function

\[
\zeta(z) = u(z^0) + \frac{p + q}{2} \cdot (z - z^0) - \frac{(q - p) \cdot (z - z^0)^2}{4\varepsilon}
\]
in $Q_{\varepsilon} := \{ z : |(q-p) \cdot (z-z^0)| \leq \varepsilon \}$. By considering separately the cases $0 \leq (q-p) \cdot (z-z^0) \leq \varepsilon$ and $-\varepsilon \leq (q-p) \cdot (z-z^0) < 0$, we find that
\[
\zeta(z) - \frac{|(q-p) \cdot (z-z^0)|^2}{4\varepsilon} = u(z^0) + \frac{p + q}{2} \cdot (z - z^0) - \frac{|(q-p) \cdot (z-z^0)|^2}{2\varepsilon} \\
\geq u(z^0) + \min \{ p \cdot (z-z^0), q \cdot (z-z^0) \} \geq u(z)
\] (6.2)
for every $z \in Q_{\varepsilon}$. Thus, $\zeta(z) - u(z) \geq 0 = \zeta(z^0) - u(z^0)$ for every $z \in Q_{\varepsilon}$. Consequently, by the definition of viscosity solution,
\[
\min \{ -A[\zeta](z_0), B(\nabla \zeta(z_0)) \} \leq 0.
\]
It is easy to calculate,
\[
\nabla \zeta(z_0) = \frac{1}{2} (p + q), \quad D^2 \zeta(z_0) = -\frac{1}{2\varepsilon} (q-p) \otimes (q-p).
\]
Since $z_0^1 \neq 0$ and $q_1 - p_1 > 0$, when $\varepsilon$ is sufficiently small, we have $-A[\zeta](z_0) > 0$. Thus, we must have $B(\frac{1}{2}(p+q)) \leq 0$. Since $1 - \mu_1 \leq p_1 < q_1 \leq 1 + \lambda_1$ we have $1 - \mu_1 < \frac{1}{2}(p_1 + q_1) < 1 + \lambda_1$. Hence, we must have one of the following:

(i) $p_2 = q_2 = 1 - \mu_2$,  \quad (ii) $p_2 = q_2 = 1 + \lambda_2$.

Let’s first consider the case (i) $p_2 = q_2 = 1 - \mu_2$. Note that $u(z_0^1, \cdot)$ is a concave function with super-differential bounded between $1 - \mu_2$ and $1 + \lambda_2$. Since $p_2 = q_2 = 1 - \mu_2$, we see that $\partial_z u(z_0^1, z_2) = 1 - \mu_2$ for all $z > z_0^0$. We define
\[
\hat{z}_0^0 = \text{inf} \{ z_2 \leq z_0^0 \mid u(z_0^1, z_2) = u(z_0^0) + (1 - \mu_2)(z_2 - z_0^0) \}.
\]
Since $u(z_0^1, z_2) \leq O(1) + (1 + \lambda_2)z_2$ for $z_2 \leq 0$, we see that $\hat{z}_0^0 > -\infty$. In addition, set $z_0^0 = (z_0^1, \hat{z}_0^0)$ we have
\[
\begin{cases}
  u(z_0^1, z_2) = u(z_0^0) + (1 - \mu_2)(z_2 - \hat{z}_0^0) & \forall z_2 \geq \hat{z}_0^0 \\
  u(z_0^1, z_2) < u(z_0^0) + (1 - \mu_2)(z_2 - \hat{z}_0^0) & \forall z_2 < \hat{z}_0^0. 
\end{cases}
\] (6.3)

From the definition of $\zeta$ and the fact that $p_2 = q_2 = 1 - \mu_2$, we obtain from (6.2) that, setting $\beta = q_1 - p_1$,
\[
\zeta(z) \geq u(z) + \frac{[\beta(z_1 - z_0^1)]^2}{4\varepsilon} \quad \forall z \in [z_0^1 - \varepsilon\beta^{-1}, z_0^1 + \varepsilon\beta^{-1}] \times \mathbb{R},
\]
\[
\zeta(z) = u(z) \quad \forall z \in \{ z_0^1 \} \times [\hat{z}_0^0, \infty),
\]
\[
\zeta(z) > u(z) \quad \forall z \in \{ z_0^1 \} \times (-\infty, \hat{z}_0^0).
\]

Using $\zeta(z_0^1, \hat{z}_0^0 - \varepsilon) > u(z_0^1, \hat{z}_0^0 - \varepsilon)$ and continuity, we can find $\eta \in (0, \varepsilon\beta^{-1})$ such that $\zeta(z_1, z_0^0 - \varepsilon) > u(z_1, z_0^0 - \varepsilon) + \eta$ for all $z_1 \in [z_0^0 - \eta, z_0^0 + \eta]$. Hence,
\[
\zeta(z_1, z_2) - u(z_1, z_2) \geq \begin{cases}
  (\beta\eta)^2/(4\varepsilon) & \text{if } |z_1 - z_0^1| = \eta, \ z_2 \in \mathbb{R} \\
  \eta & \text{if } z_2 = \hat{z}_0^0 - \varepsilon, |z_1 - z_0^1| \leq \eta.
\end{cases}
\]
Finally, set $\hat{\eta} = \min\{ (\lambda_2 + \mu_2)/2, \eta/(2\varepsilon), (\beta\eta)^2/(8\varepsilon^2) \}$ and consider the function
\[
\hat{\zeta}(z) = \zeta(z) + (z_0^0 - \hat{z}_0^0)\hat{\eta} \quad \text{in } D := (z_0^0 - \eta, z_0^0 + \eta) \times (\hat{z}_0^0 - \varepsilon, \hat{z}_0^0 + \varepsilon).
\]
Note that $\hat{\zeta}(z) > u(z)$ on the boundary $\partial D$ of $D$. In addition, $\hat{\zeta}(z_0^0) = u(z_0^0)$. Now set $m := \max_{\mathcal{D}} \{ u(z) - \hat{\zeta}(z) \}$. Then $m \geq 0$. Let $\hat{z} \in \partial \mathcal{D}$ be the point such that $u(\hat{z}) - \hat{\zeta}(\hat{z}) = m$,
then \( \hat{z} \in D \) since \( u < \hat{c} \) on \( \partial D \). Thus, \( [\hat{c}(z) + m] - u(z) \geq [\hat{c}(\hat{z}) + m] - u(\hat{z}) = 0 \) for every \( z \in D \). Hence, by definition of viscosity solution, we have

\[
\min \{-A[\hat{c} + m](\hat{z}), \ B(\nabla \hat{c}(\hat{z}))\} \leq 0.
\]

However, for every \( z \in D \),

\[
p_1 < \partial_{z_1} \hat{c}(z) < q_1, \quad \partial_{z_2} \hat{c}(z) = 1 - \mu_2 + \eta \in (1 - \mu_2, 1 - \lambda_2), \quad D^2 \hat{c}(z) = -\frac{j^2}{2\varepsilon} e_1 \otimes e_1.
\]

This implies that \( B(\nabla \hat{c}(\hat{z})) > 0 \). Hence, we must have \( -A[\hat{c}(\hat{z}) + m](\hat{z}) \leq 0 \). However, since \( \hat{z} \not\in D \), we have \( -A[\hat{c}(\hat{z}) + m](\hat{z}) \to \infty \) as \( \varepsilon \searrow 0 \); this contradicts \( -A[\hat{c}(\hat{z}) + m](\hat{z}) \leq 0 \).

Similarly, we can derive a contradiction in the second case (ii) \( p_2 = q_2 = 1 + \lambda_2 \).

The contradiction shows that \( \partial_1 u(z_1, z_2) \) is singleton if \( z_1 \neq 0 \). Now for each fixed \( z_2 \in \mathbb{R} \), consider the one dimensional function \( g(t) = u(t, z_2) \). By Lemma 4.1 (4), \( \partial g(t) = \partial_1 u(t, z_2) \) is singleton if \( t \neq 0 \). Hence, \( g \in C^1(\mathbb{R} \setminus \{0\}) \), so the classical partial derivative \( \frac{\partial u(z_1, z_2)}{\partial z_1} := g'(z_1) \) exists. Since any limit point of \( \partial_1 u(\hat{z}) \) as \( \hat{z} \to z \) is in \( \partial_1 u(z) \), we conclude that \( \lim_{z_1 \to \hat{z}_1} \frac{\partial u(z_1, z_2)}{\partial z_1} = \frac{\partial u(z_1, z_2)}{\partial z_1} \) if \( z_1 \neq 0 \). Hence, \( \frac{\partial u(z_1, z_2)}{\partial z_1} \in C(\mathbb{R}^2 \setminus \{(0) \times \mathbb{R}\}) \). As \( u \) is Lipschitz continuous, we also know that \( z_1 \partial_1 u \in C(\mathbb{R}) \). This completes the proof. \( \square \)

7. Appendix

7.1. Proof of Lemma 4.1.

(1) For each \( y \in \mathbb{R}^n \), sending \( k \to \infty \) in \( f(y) \leq f(z_k) + p_k \cdot (y - z_k) \) we obtain \( f(y) \leq f(z) + p \cdot (y - z) \), so \( p \in \partial f(z) \).

(2) Let \( \rho \) be a smooth non-negative function supported in the unit ball having unit mass. Set \( \rho_\varepsilon(z) = \rho(z/\varepsilon) \) and \( f_\varepsilon = \rho_\varepsilon * f \). Then \( f_\varepsilon \) is smooth and concave. Hence, \( \partial f_\varepsilon(z) = \{\nabla f_\varepsilon(z)\} \) where \( \nabla \) is the classical gradient. In the expression \( f_\varepsilon(y) \leq f_\varepsilon(z) + \nabla f_\varepsilon(z) \cdot (y - z) \), setting \( y = z - \nabla f_\varepsilon(z)/|\nabla f_\varepsilon(z)| \) we find that

\[
|\nabla f_\varepsilon(z)| \leq \max_{y \in B_1(z)} |f_\varepsilon(z) - f_\varepsilon(y)| \leq \max_{y, y' \in B_2(z)} |f(y_1) - f(y_2)| \quad \forall \varepsilon \in (0, 1].
\]

Let \( p \) be a limit point of \( \nabla f_\varepsilon(\hat{z}) \) as \( \varepsilon \to 0 \) and \( \hat{z} \to z \), i.e., there exists a sequence \( \{\varepsilon_j, z_j\} \) such that as \( j \to \infty \), \( (z_j, \varepsilon_j, \nabla f_\varepsilon(z_j)) \to (z, 0, p) \). Then for any \( y \in \mathbb{R}^n \), sending \( j \to \infty \) in

\[
f_{\varepsilon_j}(y) \leq f_{\varepsilon_j}(z_j) + \nabla f_{\varepsilon_j}(z_j) \cdot (y - z_j)
\]

we obtain \( f(y) \leq f(z) + p \cdot (y - z) \). This means that \( p \in \partial f(z) \), so \( \partial f(z) \) is non-empty. It is easy to show that \( \partial f(z) \) is convex and compact.

(3) Suppose \( \partial f(z) = \{p\} \) is a singleton. Then \( \lim_{\varepsilon \to 0, \varepsilon \searrow 0} \nabla f_\varepsilon(\hat{z}) = p \) since any limit point of \( \nabla f_\varepsilon(\hat{z}) \) is in \( \partial f(z) \). Hence, for any \( \beta > 0 \), there exist \( r > 0 \) and \( \varepsilon_0 > 0 \) such that \( |\nabla f_\varepsilon(\hat{z}) - p| \leq \beta \) for all \( \hat{z} \in B_r(z) \) and \( \varepsilon \in (0, \varepsilon_0] \). This implies that

\[
|f_\varepsilon(y) - f_\varepsilon(z) - p \cdot (y - z)| = \left| \int_0^1 |\nabla f_\varepsilon(z + \theta(y - x)) - p| d\theta \cdot (y - z) \right| \leq \beta |y - x|
\]

for every \( y \in B_r(z) \) and \( \varepsilon \in (0, \varepsilon_0] \). Sending \( \varepsilon \searrow 0 \) we obtain \( |f(y) - f(z) - p \cdot (y - z)| \leq |\beta |y - x| \). Thus, \( f \) is differentiable at \( z \) and \( p = \nabla f(z) \).
(4) If $\partial f(z) \neq \emptyset$, then $\partial f(z) = \{\nabla f(z)\}$ for $z$ in that neighborhood. As every limit point of $\nabla f(z)$ as $\hat{z} \to z$ is in $\partial f(z)$, we must have $\lim_{\hat{z} \to z} \nabla f(\hat{z}) = \nabla f(z)$. Hence, $f$ is $C^1$ in that open neighborhood of $z^0$.

(5) By definition, it is clear that $\partial_i f(z) \subset \partial g(0)$. We prove the reverse inclusion. Since $g(\cdot)$ is a one dimensional concave function, there are two sequences $\{t_k\}_{k=1}^{\infty}$ and $\{\tau_k\}_{k=1}^{\infty}$ with $-1/k < \tau_k < 0 < t_k < 1/k$ such that $g$ is differentiable at $\{t_k\}_{k=1}^{\infty}$ and $\partial g(0) = [a, b]$ where $a = \lim_{k \to \infty} g'(t_k)$ and $b = \lim_{k \to \infty} g'(\tau_k)$. Let $p_k \in \partial f(z + \tau_k e_1)$. Then $p_k \cdot e_1 = g'(\tau_k)$. Let $\{p_{k_j}\}$ be a subsequence such that $p_{k_j}$ converges to a limit as $k_j \to \infty$. Then $p \in \partial f(z)$. In addition, $p \cdot e_i = \lim_{j \to \infty} p_{k_j} \cdot e_i = \lim_{k \to \infty} g'(\tau_k) = b$. Similarly, one can find $q \in \partial f(z)$ such that $q \cdot e_i = \lim_{\ell \to \infty} g'(t_k) = a$. As $\partial f(z)$ is convex, we conclude that $\partial i f(z) = \{p \cdot e_i \mid p \in \partial f(z)\} \supset [a, b] = \partial g(0)$. This completes the proof.

7.2. Extension to the CRRA utility.

Now let us examine the case with the CRAR utility, namely,

$$V(c) = \frac{1}{\gamma} c^{\gamma}, \gamma < 1, \gamma \neq 0,$$

for which we require that the liquidated wealth be non-negative:

$$x + \sum_i \ell_i(y_i) \geq 0.$$

Note that we can directly consider the infinite horizon problem with the CRAR utility and the above solvency constraint. Let $\Phi(x, y_1, y_2)$ be the associated value function which satisfies (cf. Dai and Zhong (2010))

$$\min \left\{ -L \Phi, \min \{ (1 + \lambda_1) \partial_x \Phi - \partial_y \Phi, \min_i \{ (1 - \mu_i) \partial_x \Phi + \partial_y \Phi \} \right\} = 0,$$

in $x + \sum_i \ell_i(y_i) > 0$, where $L$ is as given in (3.10).

For illustration, we still consider the case of two risky assets. The homogeneity of the utility function allows us to make the following transformation:

$$z_i = \frac{y_i}{x + y_1 + y_2}, \quad i = 1, 2,$$

$$\varphi(z_1, z_2) \equiv \Phi(1 - z_1 - z_2, z_1, z_2) = \frac{1}{(x + y_1 + y_2)^\gamma} \Phi(x, y_1, y_2).$$

Then (7.1) reduces to

$$\min \left\{ -\hat{A} \varphi, \min_i \left\{ \lambda_i \gamma \varphi - \sum_k (\delta_{ik} + \lambda_i z_k) \partial_{z_k} \varphi \right\}, \min_i \left\{ \mu_i \gamma \varphi - \sum_k (-\delta_{ik} + \mu_i z_k) \partial_{z_k} \varphi \right\} \right\} = 0$$

in $D = \{(z_1, z_2) : 1 - \sum_i |z_i - \ell_i(z_i)| > 0\}$, where the expression of $\hat{A}$ is omitted.

Now we define

$$B_i = \left\{ (z_1, z_2) \in D : \lambda_i \gamma \varphi - \sum_k (\delta_{ik} + \lambda_i z_k) \partial_{z_k} \varphi = 0 \right\},$$

$$S_i = \left\{ (z_1, z_2) \in D : \mu_i \gamma \varphi - \sum_k (-\delta_{ik} + \mu_i z_k) \partial_{z_k} \varphi = 0 \right\},$$

$$N_i = D \cap B_i^c \cap S_i^c.$$
We aim to show that $D$ is partitioned into nine regions as shown in Figure 2. In particular, $N_1 \cap N_2$ has four distinct corners. We point out that the shape of trading/no-trading regions is the same as that postulated by Bichuch and Shreve (2011) who deal with a slightly different setting: the prices of risky assets follow \textit{arithmetic} Brownian motions.

Let us take as an example the region $S_1 \cap S_2$. Recall the proof for the CARA utility case in which two conditions play critical roles: 1) $\varphi$ is concave and is $C^1$ except on the coordinates planes; 2) $\partial_z \varphi$, $i = 1, 2$ are constants in $S_1 \cap S_2$. The former still holds true with the CRRA utility whereas the latter does not. This motivates us to make a new transformation:

$$
\tau_i \equiv \frac{y_i}{x + (1 - \mu_1) y_1 + (1 - \mu_2) y_2}, \quad i = 1, 2,
$$

$$
\overline{\varphi} (\overline{\tau}_1, \overline{\tau}_2) \equiv \Phi (1 - (1 - \mu_1) \overline{\tau}_1 - (1 - \mu_2) \overline{\tau}_2, \overline{\tau}_1, \overline{\tau}_2)
$$

$$
= \frac{1}{[x + (1 - \mu_1) y_1 + (1 - \mu_2) y_2]^{\gamma}} \Phi (x, y_1, y_2).
$$

It is easy to verify that

\[-(1 - \mu_i) \partial_{z_i} \Phi + \partial_{y_i} \Phi = [x + (1 - \mu_1) y_1 + (1 - \mu_2) y_2]^{\gamma - 1} \partial_{z_i} \overline{\varphi},\]

Figure 2. Shape of trading and no-trading regions with CRRA utility
which implies

$$\partial z_i \varphi = 0 \text{ in } S_1 \cap S_2, \ i = 1, 2.$$  

Since $\varphi$ is also concave and is $C^1$ except on the coordinates planes, we can use the same analysis developed in the CARA utility case to obtain the desired result. We point out that most results derived earlier with the CARA utility can be extended to the CRRA utility case.

References


