4

Consumption–Investment Problems

4.1 Consumption–Investment without Friction

4.1.1 The Merton Problem

The study of consumption–investment problems in continuous time was initiated by Merton. He considered a model of frictionless market where the price processes are geometric Brownian motions and the investor's goal is to maximize the expected discounted utility of consumption on the infinite time interval. For the power utility function, he obtained an explicit solution of the optimal control problem. This solution has a clear financial meaning: the optimal investment is to keep the proportions of the total wealth held in risky securities equal to a constant vector. The latter is easily calculated from the model parameters. This work was extended by many authors in various directions including models with transaction costs, which are the main objects of our interest. Taking into account that the Merton problem is classical and exposed in a number of textbooks, we give here a rather sketchy presentation needed to understand basic ideas and methods as well as their evolution. The results of this section will be used at the end of this chapter, where we discuss an asymptotical behavior of the consumption–investment problem for small transaction cost coefficients.

We are given a stochastic basis with an m-dimensional standard Wiener process w. The market contains a nonrisky security, which is the *numéraire*, i.e., its price is identically equal to unit, and m risky securities with the price evolution

$$dS_t^i = S_t^i\big(\mu^i\,dt + dM_t^i\big), \quad i = 1, \dots, m, \tag{4.1.1}$$

where $M = \Sigma w$ is a (deterministic) linear transform of w. Thus, M is a Gaussian martingale with $\langle M\rangle_t = At$; the covariance matrix $A = \Sigma\Sigma^*$ is assumed to be nondegenerate.

The evolution of the value process corresponding to a self-financing strategy H is given as $dV_t = H_t\,dS_t$. Assuming withdrawal of the funds for the

Yu. Kabanov, M. Safarian, *Markets with Transaction Costs*,
Springer Finance, DOI 10.1007/978-3-540-68121-2_4,

consumption with rate $c_t \geq 0$, we arrive at the dynamics

$$dV_t = H_t\, dS_t - c_t\, dt. \tag{4.1.2}$$

Of course, we can substitute dS_t by its expression given in (4.1.1). Since H_t^i is a number of units of the ith asset in the portfolio, the quantity $\alpha_t^i := H_t^i S_t^i / V_t$ can be interpreted as the proportion of the wealth invested in this asset. It is convenient to choose α together with c as the control parameters. With these considerations, the problem with infinite time horizon can be formulated in usual terms of stochastic optimal control theory in the following way.

The system dynamics is given by the controlled stochastic differential equation

$$dV_t = V_t \alpha_t (\mu\, dt + dM_t) - c_t\, dt, \qquad V_0 = x, \tag{4.1.3}$$

with the initial condition $x > 0$ and the control $\pi = (\alpha, c)$, which is a predictable process. We suppose that the consumption intensity process c has trajectories integrable on every finite interval, while the trajectories of α are uniformly bounded by a constant which may depend on the strategy.[1]

More substantially, we require from π to be in the class of *admissible controls* $\mathcal{A}(x)$ for which the process $V = V^{x,\pi}$ is positive. We assume also that after the bankruptcy time (which is the first instant when V hits zero), the control π is equal to zero, and the process V stops.

The investor's goal is the following:

$$EJ_\infty^\pi \to \max, \tag{4.1.4}$$

i.e., to maximize the expectation of the limiting expected value of the utility process J^π defined as

$$J_t^\pi := \int_0^t e^{-\beta s} u(c_s)\, ds. \tag{4.1.5}$$

The standard economically meaningful assumptions on the utility function are that u is increasing and concave. For the sake of simplicity, we add to this that u is positive and $u(0) = 0$. The parameter $\beta > 0$ shows to which extent the agent prefers to consume today rather than in the future.

A typical example is the power utility function $u(c) = c^\gamma/\gamma$, $\gamma \in]0,1[$.

Define the Bellman function

$$W(x) := \sup_{\pi \in \mathcal{A}(x)} EJ_\infty^\pi, \quad x > 0. \tag{4.1.6}$$

By convention, $\mathcal{A}(0) := \{0\}$ and $W(0) := 0$.

Notice that the Bellman function W inherits the properties of u. Namely, it is increasing (as $\mathcal{A}(\tilde{x}) \supseteq \mathcal{A}(x)$ when $\tilde{x} \geq x$). With the chosen α-parameterization, its concavity appears not to be so obvious, but we get it immediately

[1] This assumption is not very wise but allows us to avoid discussions of integrability. It is done because, in the Merton problem, the optimal strategy in a wider class possesses this property.

by turning back to the initial H-parameterization. Indeed, suppose that the strategies $\pi_j = (\alpha_j, c_j)$, $\pi_j \in \mathcal{A}(x_j)$, $j = 1, 2$, generate the value processes V_j. The convex combination of these processes, $V = \lambda V_1 + (1-\lambda)V_2$, is of the form (4.1.2), where H and c are the convex combinations with the same coefficients of the corresponding controls. The process H admits the representation via the process α with the components

$$\alpha^i = H^i S^i / V = \frac{\lambda V_1}{\lambda V_1 + (1-\lambda)V_2}\alpha_1^i + \frac{(1-\lambda)V_2}{\lambda V_1 + (1-\lambda)V_2}\alpha_2^i;$$

α is bounded because both α_j are bounded. Thus, $\pi = (\alpha, \lambda c_1 + (1-\lambda)c_2)$ belongs to $\mathcal{A}(x)$ with $x = \lambda x_1 + (1-\lambda)x_2$, and, therefore,

$$W\big(\lambda x_1 + (1-\lambda)x_2\big) \geq EJ_\infty^\pi \geq \lambda EJ_\infty^{\pi_1} + (1-\lambda)EJ_\infty^{\pi_2}$$

due to the concavity of u. With this, we obtain the concavity of W by taking supremum over π_1 and π_2.

Notice that we cannot guarantee without additional assumptions that W is finite. If the latter property holds, then, due to the concavity, $W(x)$ is continuous for $x > 0$, but the question whether it is continuous at zero remains open.

At last, when the utility u is a power function, the Bellman function W, if finite, is proportional to u. Indeed, the linear dynamics of the control system implies that $W(\nu x) = \nu^\gamma W(x)$ for all $\nu > 0$, i.e., the Bellman function is positive homogeneous of the same order as the utility function. In a scalar case this homotheticity property defines, up to a multiplicative constant, a unique finite function, namely x^γ.

Now we formulate the Merton theorem.

Theorem 4.1.1 *Let u be the power utility function. Assume that the parameters of the model are such that the constant*

$$\kappa_M := \frac{1}{1-\gamma}\left(\beta - \frac{1}{2}\frac{\gamma}{1-\gamma}\left|A^{-1/2}\mu\right|^2\right) > 0. \tag{4.1.7}$$

Then the optimal strategy $\pi^o = (\alpha^o, c^o)$ is given by the formulae

$$\alpha^o = \theta := \frac{1}{1-\gamma}A^{-1}\mu, \tag{4.1.8}$$

$$c_t^o = \kappa_M V_t^o, \tag{4.1.9}$$

where V^o is the solution of the linear stochastic equation

$$dV^o = V_t^o \theta(\mu\, dt + dM_t) - \kappa_M V_t^o\, dt, \qquad V_0^o = x. \tag{4.1.10}$$

The process V^o is optimal, and the Bellman function is

$$W(x) = \kappa_M^{\gamma-1} x^\gamma / \gamma = \mathbf{m} x^\gamma. \tag{4.1.11}$$

Note that W is proportional to x^γ, and, therefore, the last assertion is about the exact value of the coefficient $\mathbf{m}$, which happens to be finite and equal to $\kappa_M^{\gamma-1}/\gamma$ with κ_M given by (4.1.7).

4.1.2 The HJB Equation and a Verification Theorem

The most powerful and efficient method to solve stochastic control problems is the method of dynamic programming based on the analysis of the Hamilton–Jacobi–Bellman equation (HJB, in short). For our infinite-horizon problem, the latter is

$$\sup_{(\alpha,c)}\left[\frac{1}{2}\left|A^{1/2}\alpha\right|^2 x^2 f''(x)+\alpha\mu x f'(x)-\beta f(x)-f'(x)c+u(c)\right]=0, \quad (4.1.12)$$

where $x>0$, and the supremum is taken over all $\alpha\in\mathbf{R}^d$ and $c\in\mathbf{R}_+$.

To solve the consumption–investment problem with the power utility function, we use a very elementary tool, namely, the so-called verification theorem for the HJB equation. It is based on the following considerations.

Let $f:\mathbf{R}_+\to\mathbf{R}_+$ and $\pi\in\mathcal{A}(x)$. We consider the nonnegative process $X^f=X^{f,x,\pi}$ with

$$X_t^f:=e^{-\beta t}f(V_t)+J_t^\pi, \quad (4.1.13)$$

where $V=V^{x,\pi}$. If f is smooth, the Itô formula for the process V given by (4.1.3) implies the following important representation, which is the key point to explain how the HJB equation arises:

$$X_t^f=f(x)+D_t+N_s, \quad (4.1.14)$$

where

$$D_t:=\int_0^t e^{-\beta s}L(V_s,\alpha_s,c_s)\,ds \quad (4.1.15)$$

with $L(x,\alpha,c)$ standing for the expression in square brackets of the formula (4.1.12), and

$$N_t:=\int_0^t e^{-\beta s}f'(V_s)V_s\alpha_s\,dM_s. \quad (4.1.16)$$

The process N is a continuous local martingale up to the bankruptcy time σ. That is, there exist stopping times $\sigma_n\uparrow\sigma$ such that the stopped processes N^{σ_n} are uniformly integrable martingales. In the case where $\sigma=\infty$ and N is a martingale, we shall take $\sigma_n=n$.

Suppose now that a smooth function f is a supersolution of (4.1.12), i.e.,

$$\sup_{(\alpha,c)}\left[\frac{1}{2}\left|A^{1/2}\alpha\right|^2 x^2 f''(x)+\alpha\mu x f'(x)-\beta f(x)-f'(x)c+u(c)\right]\le 0. \quad (4.1.17)$$

Then the integrand in the definition of D does not exceed zero, and, therefore, the process D is decreasing with $D_0=0$. This implies, in particular, the inequality $N\ge -f(x)$. It follows (as usual, by applying the Fatou lemma) that N, being bounded from below, is a supermartingale. Due to the inequality $-D_t\le f(x)+N_t$, the (negative) random variable D_t is integrable: we obtain that the process X_t^f is a supermartingale, and, hence,

$$EJ_t = EX_t^f - Ee^{-\beta t} f(V_t) \le EX_t^f \le f(x). \tag{4.1.18}$$

Since $EJ_t \to EJ_\infty$ as $t \to \infty$, we infer that $W(x) \le f(x)$, i.e., f provides a "cap" for the Bellman function, implying, in particular, that the latter is finite. If, moreover, the supersolution f vanishes at zero, the function W (being positive) is necessarily continuous at zero. Summarizing, we formulate the outcome of this reasoning in the following statement.

Proposition 4.1.2 *If f is a supersolution of (4.1.12), then $W \le f$, and, hence, $W \in C(\mathbf{R}_+ \setminus \{0\})$. If, moreover, $f(0+) = 0$, then $W \in C(\mathbf{R}_+)$.*

An inspection of the above reasoning shows that if, in addition, it happened that the process D (depending on the control) vanishes and

$$\lim_n Ee^{-\beta\sigma_n} f(V_{\sigma_n}) = 0, \tag{4.1.19}$$

then $W = f$, and the corresponding control is optimal. With these observations, we arrive at the promised verification theorem, which can be obtained, of course, in a much more general context.

Theorem 4.1.3 *Let $f \in C(\mathbf{R}_+) \cap C^2(\mathbf{R}_+ \setminus \{0\})$ be a positive concave function solving the HJB equation (4.1.12) and vanishing at zero. Suppose that the supremum in (4.1.12) is attained on $\alpha(x)$ and $c(x)$ such that α is a bounded measurable function, c is a positive measurable function, and the equation*

$$dV_t^o = V_t^o \alpha\big(V_t^o\big)(\mu\, dt + dM_t) - c\big(V_t^o\big)\, dt, \qquad V_0^o = x, \tag{4.1.20}$$

admits a strong solution V_t^o. If condition (4.1.19) holds for the process V^o, then $W = f$, and the optimal control $\pi^o = (\alpha(V^o), c(V^o))$.

4.1.3 Proof of the Merton Theorem

With the above provision, we return to the HJB equation (4.1.12) and calculate the supremum.

Put

$$u^*(p) := \sup_{c \ge 0} \big[u(c) - cp\big];$$

the function u^* is the Fenchel transform of the function $-u(-.)$. In particular, for the power utility $u(c) = c^\gamma/\gamma$, we have that

$$u^*(p) = \frac{1-\gamma}{\gamma} p^{\gamma/(\gamma-1)} \tag{4.1.21}$$

because the supremum in the definition of u^* is attained at the point $p^{1/(\gamma-1)}$. Expecting that $f'' < 0$, we find easily that the maximum of the quadratic form over α is attained at the point

$$\alpha^o(x) = -A^{-1}\mu\frac{f'(x)}{xf''(x)}.$$

Thus, the HJB equation can be transformed to the following one:

$$-\frac{1}{2}\big|A^{-1/2}\mu\big|^2\frac{(f'(x))^2}{f''(x)} - \beta f(x) + \frac{1-\gamma}{\gamma}\big(f'(x)\big)^{\frac{\gamma}{\gamma-1}} = 0.$$

We find easily that its solution of the form $f(x) = \mathbf{m}x^\gamma$ should have the coefficient $\mathbf{m} = \kappa_M^{\gamma-1}/\gamma$ with $\kappa_M > 0$ given in (4.1.7).

Now the function $\alpha^o(x) = A^{-1}\mu/(1-\gamma)$ is constant, $c^o(x) = \kappa_M x$, and (4.1.20), pretending to describe the optimal dynamics, is linear:

$$\frac{dV_t^o}{V_t^o} = \left(\frac{1}{1-\gamma}\big|A^{-1/2}\mu\big|^2 - \kappa_M\right)dt + \frac{A^{-1}\mu}{1-\gamma}\,dM, \qquad V_0^o = x,$$

and its solution is the geometric Brownian motion, which never hits zero. Noticing that $\langle A^{-1}\mu M\rangle_t = |A^{-1/2}\mu|^2 t$, it can be given by the following explicit formula:

$$V_t^o = x\exp\left\{\left(\frac{1}{1-\gamma} - \frac{1}{2}\frac{1}{(1-\gamma)^2}\right)\big|A^{-1/2}\mu\big|^2 t - \kappa_M t + \frac{A^{-1}\mu}{1-\gamma}M_t\right\}.$$

Since $E(V_t^o)^p = x^p e^{\kappa_p t}$ where κ_p is a constant, the process N for this control is a true martingale, and we may take the localizing sequence σ_n deterministic.

In the particular case where $p = \gamma$, the corresponding constant

$$\kappa_\gamma = \frac{1}{2}\frac{\gamma}{1-\gamma} - \gamma\kappa_M = \beta - \kappa_M$$

in virtue of (4.1.7). Thus,

$$e^{-\beta t}E\big(V_t^o\big)^\gamma = x^\gamma e^{-\kappa_M t},$$

and, therefore, (4.1.19) holds. The Merton theorem is proven.

4.1.4 Discussion

1. The optimal strategy in the Merton problem with the power utility functions prescribes to keep constant proportions of wealth in each position. Let us consider the special case $m = 1$, i.e., the model with a single risky asset. Then the quantities $V_t^{2o} := \alpha^o V_t^o$ and $V_t^{1o} = (1-\alpha^o)V_t^o$ are, respectively, the optimal holdings in the risky and nonrisky assets,

$$\alpha^o = \theta = \frac{1}{1-\gamma}\frac{\mu}{\sigma^2}.$$

Thus,

$$V_t^{2o} := \frac{\alpha^o}{1-\alpha^o} V_t^{1o} = \frac{\theta}{1-\theta} V_t^{1o}.$$

This means that the two-dimensional process (V_t^{1o}, V_t^{2o}) on the plain (v^1, v^2) evolves along the straight line with the slope $\theta/(1-\theta)$, called in the literature the *Merton line*. The parameter θ is referred to as the *Merton proportion*.

2. In our presentation we consider the case where the price of the nonrisky asset is constant over time as it would pay the interest $r = 0$. The reader may be accustomed with the tradition to treat the model with an arbitrary $r \geq 0$. However, it is easy to see that, for the power utility function, considering the model with zero interest rate does not lead to any loss in generality. Indeed, due to the identity

$$u\big(e^{rs} c_s\big) = e^{\gamma r s} u(c_s),$$

the maximization problem where the consumption is measured in "money" is the same as that where the consumption is measured in "bonds" but with the coefficient β replaced by $\tilde{\beta} := \beta - \gamma r$. Thus, there is no real reason to retain r in calculations.

3. An analysis of the proof of Theorem 4.1.1 shows that, after minor changes, it works well also for the power utility function with $\gamma < 0$, and, hence, the same explicit formulae represent the optimal solution also in this case. The HJB approach can be extended to the model with the logarithmic utility function $u(c) = \ln c$ (corresponding to the value $\gamma = 0$). Of course, one needs to impose an additional constraint to the consumption process ensuring the integrability of J_∞^π.

4. Turning back to the multi-asset case, let us define the scalar process $\tilde{M}$ with $d\tilde{M} = \theta(\mu\, dt + dM_t)$. Let us consider the same consumption–investment problem imposing the restriction that the investments should be shared between money and the risky asset the price evolution of which follows the process $\tilde{M}$. Any value process and consumption process in this two-asset model are those of the original one. One can imagine a financial institution (a mutual fund) which offers such an artificial asset, called the *market portfolio*. This allows the agent to allocate his wealth only in the nonrisky asset and the market portfolio. Due to this economical interpretation, the Merton theorem sometimes is referred to as the *mutual fund theorem*.

5. Formula (4.1.11) shows that, for a positive initial capital, the value $W(x) \to \infty$ as $\kappa_M \downarrow 0$. It follows that, for small values of the discount parameter β, namely, when

$$\beta \leq \frac{1}{2} \frac{\gamma}{1-\gamma} \big|A^{-1/2}\mu\big|^2,$$

the Bellman function $W(x) = \infty$, $x > 0$.

4.1.5 Robustness of the Merton Solution

There is an interesting question about the sensitivity of the Merton solution with respect to errors in determining the optimal proportion. It happens that it is quite robust: a deviation of order ε from the Merton proportion leads to losses in the expected utility only of order ε^2. To see this, suppose that in the two-asset model the investor's strategy is to maintain the proportion $\alpha^o+\varepsilon$ and consume a constant part $(1+\delta)\kappa_M$ of the current wealth optimizing the expected utility with respect to δ. Assume, for simplicity, that the initial endowment $x = 1$. For such a strategy, the dynamics is given by the linear equation

$$\frac{dV_t}{V_t} = \big(\alpha^o + \varepsilon\big)(\mu\, dt + \sigma\, dw_t) - (1+\delta)\kappa_M\, dt$$

the solution of which is the geometric Brownian motion

$$V_t = \exp\Big\{ \big(\alpha^o + \varepsilon\big)\mu t - \frac{1}{2}\big(\alpha^o + \varepsilon\big)^2 \sigma^2 t - (1+\delta)\kappa_M t + \big(\alpha^o + \varepsilon\big)\sigma w_t \Big\}.$$

We have that

$$EV_t^\gamma = e^{\kappa_\gamma(\varepsilon,\delta)t},$$

where

$$\kappa_\gamma(\varepsilon,\delta) = \beta - \kappa_M - \frac{1}{2}\gamma(1-\gamma)\sigma^2\varepsilon^2 - \gamma\kappa_M\delta,$$

and, in particular, $\kappa_\gamma(0,0) = \kappa_\gamma = \beta - \kappa_M$.

Notice that the coefficient at ε is zero, and this is a crucial fact. It follows that

$$EJ_\infty = \frac{1}{\gamma}\kappa_M^\gamma(1+\delta)^\gamma \int_0^\infty e^{-\beta t} EV_t^\gamma\, dt = \frac{1}{\gamma}\kappa_M^{\gamma-1} \frac{(1+\delta)^\gamma}{1 + \frac{1}{2\kappa_M}\gamma(1-\gamma)\sigma^2\varepsilon^2 + \gamma\delta}.$$

Maximization over δ gives us the optimal value $\delta^o = \frac{1}{2\kappa_M}\gamma\sigma^2\varepsilon^2$, for which

$$EJ_\infty = \frac{1}{\gamma}\kappa_M^{\gamma-1}\big(1+\delta^o\big)^{\gamma-1} = \mathbf{m} - \frac{1}{2}(1-\gamma)\kappa_M^{\gamma-2}\sigma^2\,\varepsilon^2 + O\big(\varepsilon^4\big),$$

and we get the claimed asymptotic.

Of course, the robustness of the Merton solution is of great practical importance.

4.2 Consumption–Investment under Transaction Costs

4.2.1 The Model

The setting described in this section is, in some aspects, slightly more general than that of the standard model of financial market under constant proportional transaction costs. In particular, the cone K is not supposed to be polyhedral. On the other hand, it is more restrictive with respect to the price processes: they are assumed to be geometric Brownian motions. Our framework appeals to a well-developed theory of viscosity solutions (in fact, only to basic elements of the latter) and allows us to catch essential properties of the Bellman function before going to the specific case of the two-asset model with the power utility function the detailed analysis of which is our ultimate goal.

Let $Y = (Y_t)$ be an $\mathbf{R}^d$-valued semimartingale on a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with trivial initial σ-algebra. Let K and $\mathcal{C}$ be *proper* cones in $\mathbf{R}^d$ such that $\mathcal{C} \subseteq \operatorname{int} K \neq \emptyset$. Define the set $\mathcal{A}$ of controls $\pi = (B, C)$ as the set of adapted càdlàg processes of bounded variation such that, up to an evanescent set,

$$\dot{B} \in -K, \qquad \dot{C} \in \mathcal{C}. \tag{4.2.1}$$

Let $\mathcal{A}_a$ be the set of controls with absolutely continuous C and $\Delta C_0 = 0$. For the elements of $\mathcal{A}_a$, we have $c := dC/dt \in \mathcal{C}$.

The controlled process $V = V^{x,\pi}$ is the solution of the linear system

$$dV_t^i = V_{t-}^i \, dY_t^i + dB_t^i - dC_t^i, \quad V_{0-}^i = x^i, \quad i = 1, \dots, d. \tag{4.2.2}$$

For $x \in \operatorname{int} K$, we consider the subsets $\mathcal{A}^x$ and $\mathcal{A}_a^x$ of "admissible" controls for which the processes $V^{x,\pi}$ never leave the set $\operatorname{int} K \cup \{0\}$ and have the origin as an absorbing point. Thus, if $V_{s-}(\omega) \in \partial K$, then $\Delta B_s(\omega) = -V_{s-}(\omega)$.

The important hypothesis that the cone K is proper, i.e., $K \cap (-K) = \{0\}$, or equivalently, $\operatorname{int} K^* \neq \emptyset$, corresponds to the model of financial market with *efficient friction*. In a financial context, K (usually containing $\mathbf{R}_+^d$) is interpreted as the solvency region and $C = (C_t)$ as the consumption process; the process $B = (B_t)$ describes the accumulated fund transfers.

Let $G := (-K) \cap \partial\mathcal{O}_1(0)$, where $\partial\mathcal{O}_1(0) = \{x \in \mathbf{R}^d : \ |x| = 1\}$ in accordance with the notation for the open ball $\mathcal{O}_r(y) := \{x \in \mathbf{R}^d : \ |x - y| < r\}$ The set G is a compact, and $-K = \operatorname{cone} G$. We denote by Σ_G the *support function* of G, given by the relation $\Sigma_G(p) = \sup_{x \in G} px$.

We shall work using the following assumption:

$\mathbf{H_1}$. The process Y is a continuous process with independent increments with mean $EY_t = \mu t$, $\mu \in \mathbf{R}^d$, and covariance $DY_t = At$.

To facilitate references, we formulate also a more specific hypothesis (frequent in the literature), where the matrix A is diagonal with $a^{ii} = (\sigma^i)^2$, i.e., the components of the driving noise are independent.

$\mathbf{H_2}$. The components of Y are of the form $dY_t^i = \mu^i \, dt + \sigma^i \, dw_t^i$, where w is a standard Wiener process in $\mathbf{R}^d$.

In our proof of the dynamic programming principle (needed to derive the HJB equation) we shall assume that the stochastic basis is a canonical one, that is, the space of continuous functions with the Wiener measure.

The efficient friction assumption, together with the hypothesis $\mathbf{H_1}$, ensures that the L^2-norm of the "maximal function" of the portfolio trajectories admits an exponential bound which is uniform with respect to strategies. This result will be used in the sequel to claim that certain stochastic integrals are not just local martingales but true martingales. For future references, we immediately give a precise formulation and proof.

Proposition 4.2.1 *There is a constant* $\kappa > 0$ *such that*

$$E \sup_{t \le T} |V_t|^2 \le \kappa |x|^2 e^{\kappa T^2} \tag{4.2.3}$$

for any value process $V = V^{x,\pi}$, $x \in \operatorname{int} K$, *and* $T \ge 0$.

Proof. As usual, κ denotes a "generic" positive constant which may be different in different formulae. Let us take an arbitrary vector $p \in \operatorname{int} K^*$ with $|p| = 1$. Making use that $p\,dB \le 0$ and $p\,dC \ge 0$ (in the sense of densities), we obtain from (4.2.2) that

$$pV_s \le px + \int_0^s \tilde{p} V_r \, dr + \int_0^s V_r \, d\tilde{M}_r,$$

where $\tilde{p}^i := p^i \mu^i$, and $\tilde{M}^i = p^i M^i$ with M denoting the martingale part of Y. The crucial observation is that there is $\kappa > 0$ such that $\kappa^{-1}|y| \le py$ for any $y \in K$. Since $|py| \le |y|$ for any $y \in \mathbf{R}^d$, we easily obtain the estimate

$$|V_s| \le \kappa |x| + \kappa \int_0^s |V_r| \, dr + \kappa \left| \int_0^s V_r \, d\tilde{M}_r \right|.$$

Notice that the right-hand side of this inequality is a continuous process, and, hence, V is locally bounded, i.e., there exists a sequence of stopping times $\tau_n \uparrow \infty$ such that each stopped process $V^{\tau_n} = (V_{t \wedge \tau_n})$ is bounded. With this observation, the proof is completed by a fairly standard argument, which we only sketch on. Squaring the above inequality, we get, by elementary estimates combined with the Cauchy–Schwarz and Doob inequalities, that the (bounded) function $\varphi_t^{(n)} := E \sup_{s \le t \wedge \tau_n} |V_s|^2$ satisfies the inequality

$$\varphi_t^{(n)} \le \kappa |x|^2 + \kappa (T+1) \int_0^t \varphi_s^{(n)} \, ds.$$

The Gronwall–Bellman lemma implies that

$$\varphi_T^{(n)} \le \kappa |x|^2 e^{\kappa (T+1) T}.$$

Taking here the limit in n and enlarging the constant, we arrive at the required bound. □

4.2.2 Goal Functionals

Let $U : \mathcal{C} \to \mathbf{R}_+$ be a concave function such that $U(0) = 0$ and $U(x)/|x| \to 0$ as $|x| \to \infty$. With every $\pi = (B, C) \in \mathcal{A}_a^x$, we associate the "utility process"

$$J_t^\pi := \int_0^t e^{-\beta s} U(c_s)\, ds, \quad t \geq 0,$$

where $\beta > 0$. We consider the infinite-horizon maximization problem with the *goal functional* EJ_∞^π and define its *Bellman function* W by

$$W(x) := \sup_{\pi \in \mathcal{A}_a^x} EJ_\infty^\pi, \quad x \in \operatorname{int} K. \tag{4.2.4}$$

If π_i, $i = 1, 2$, are admissible strategies for the initial points x_i, then the strategy $\lambda\pi_1 + (1 - \lambda)\pi_2$ is an admissible strategy for the initial point $\lambda x_1 + (1-\lambda)x_2$ for any $\lambda \in [0, 1]$, and the corresponding absorbing time is the maximum of the absorbing times for both π_i. It follows that the function W is *concave* on $\operatorname{int} K$. Since $A_a^{x_1} \subseteq A_a^{x_2}$ when $x_2 - x_1 \in K$, the function W is *increasing* with respect to the partial ordering $\geq_K$ generated by the cone K. It is convenient to put W equal to zero on the boundary of K and extend it to the whole space $\mathbf{R}^d$ as a concave function just by putting $W := -\infty$ outside K.

Remark 1. In financial models, usually, $\mathcal{C} = \mathbf{R}_+ e_1$ and $\sigma^0 = 0$, i.e., the only first (nonrisky) asset is consumed. Our presentation in this section is oriented to the scalar power utility function $u(c) = c^\gamma/\gamma$, $\gamma \in]0, 1[$. As we already mentioned in the previous section, in this case there is no need to consider a nonzero interest rate for the nonrisky asset, which can be chosen as the numéraire. Of course, for other types of utility functions, adding to the model an interest rate may have sense.

Remark 2. We consider here a model with mixed "regular–singular" controls. In fact, the assumption that the consumption process has an intensity $c = (c_t)$ and the agent's utility depends on this intensity is not very satisfactory from the economical point of view. One can consider models with an intertemporal substitution and the consumption by "gulps," i.e., dealing with "singular" controls of the class $\mathcal{A}^x$ and the goal functionals like

$$J_t^\pi := \int_0^t e^{-\beta s} U(\bar{C}_s)\, ds,$$

where

$$\bar{C}_s = \int_0^s K(s, r)\, dC_r$$

with a suitable kernel $K(s, r)$ (the exponential kernel $e^{-\gamma(s-r)}$ is the common choice).

4.2.3 The Hamilton–Jacobi–Bellman Equation

Assume that hypothesis $\mathbf{H}_1$ on the structure of driving noise holds. In the sequel we denote by U^* the convex function $U^*(p) := \sup_{x\in\mathcal{C}}(U(x) - px)$. We introduce a continuous function of four variables by putting

$$F(X, p, W, x) := \max\{F_0(X, p, W, x) + U^*(p), \Sigma_G(p)\},$$

where X belongs to $\mathcal{S}_d$, the set of $d\times d$ symmetric matrices, $p, x \in \mathbf{R}^d$, $W \in \mathbf{R}$, and the function F_0 is given by

$$F_0(X, p, W, x) := \frac{1}{2}\mathrm{tr}\, A(x)X + \mu(x)p - \beta W,$$

where $A^{ij}(x) := a^{ij}x^i x^j$, $\mu^i(x) := \mu^i x^i$, $1 \le i, j \le d$. In the detailed form we have that

$$F_0(X, p, W, x) = \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}x^i x^j X^{ij} + \sum_{i=1}^{d} \mu^i x^i p^i - \beta W.$$

If ϕ is a smooth function, we put

$$\mathcal{L}\phi(x) := F\big(\phi''(x), \phi'(x), \phi(x), x\big).$$

In a similar way, $\mathcal{L}_0$ corresponds to the function F_0.

We show, under mild hypotheses, that W is the unique viscosity solution of the Dirichlet problem for the HJB equation

$$F\big(W''(x), W'(x), W(x), x\big) = 0, \quad x \in \mathrm{int}\, K, \tag{4.2.5}$$

$$W(x) = 0, \quad x \in \partial K, \tag{4.2.6}$$

with the boundary condition understood in the usual classical sense.

We do not suppose that the reader is acquainted with the theory of viscosity solutions. Necessary prerequisites, adapted to our needs, are given in the next sections.

4.2.4 Viscosity Solutions

Since, in general, W may have no derivatives at some points $x \in \mathrm{int} K$ (and this is, indeed, the case for the model considered here), the notation (4.2.5) needs to be interpreted. The idea of viscosity solutions is to plug into F the derivatives and Hessians of quadratic functions touching W from above and below. Formal definitions (adapted to the case we are interested in) are as follows.

Let f and g be functions defined in a neighborhood of zero. We shall write $f(.) \lessapprox g(.)$ if $f(h) \le g(h) + o(|h|^2)$ as $|h| \to 0$. The notation $f(.) \gtrapprox g(.)$ and $f(.) \approx g(.)$ has an obvious meaning.

For $p \in \mathbf{R}^d$ and $X \in \mathcal{S}_d$, we consider the quadratic function

$$Q_{p,X}(z) := pz + (1/2)\langle Xz, z\rangle, \quad z \in \mathbf{R}^d,$$

and define the *superjets* and *subjets* of a function v at the point x:

$$J^+v(x) := \big\{(p,X): \ v(x+.) \lessapprox v(x) + Q_{p,X}(.)\big\},$$
$$J^-v(x) := \big\{(p,X): \ v(x+.) \gtrapprox v(x) + Q_{p,X}(.)\big\}.$$

In other words, $J^+v(x)$ (resp. $J^-v(x)$) is the family of coefficients of quadratic functions $v(x) + Q_{p,X}(y - .)$ dominating the function $v(.)$ (resp., dominated by this function) in a neighborhood of the point x with precision up to the second order included and coinciding with $v(.)$ at this point.

A function $v \in C(K)$ is called a *viscosity supersolution* of (4.2.5) if

$$F\big(X, p, v(x), x\big) \le 0 \quad \forall (p,X) \in J^-v(x), \ x \in \operatorname{int} K.$$

A function $v \in C(K)$ is called a *viscosity subsolution* of (4.2.5) if

$$F\big(X, p, v(x), x\big) \ge 0 \quad \forall (p,X) \in J^+v(x), \ x \in \operatorname{int} K.$$

A function $v \in C(K)$ is a *viscosity solution* of (4.2.5) if v is simultaneously a viscosity supersolution and subsolution of (4.2.5).

At last, a function $v \in C(K)$ is called a *classical supersolution* of (4.2.5) if $v \in C^2(\operatorname{int} K)$ and $\mathcal{L}v \le 0$ on $\operatorname{int} K$. We add the adjective *strict* when $\mathcal{L}v < 0$ on the set $\operatorname{int} K$.

Of course, the above notions[2] can be formulated also for open subsets of K.

If v is smooth at a point x, then

$$J^+v(x) := \big\{(p,X): \ p = v'(x), X \ge v''(x)\big\},$$
$$J^-v(x) := \big\{(p,X): \ p = v'(x), X \le v''(x)\big\},$$

where the inequality between matrices is understood in the sense of partial ordering induced by the cone of positive semidefinite matrices. The pair $(v'(x), v''(x))$ is the unique element belonging to the intersection of $J^-v(x)$ and $J^+v(x)$. Thus, any viscosity solution v which is in $C^2(\operatorname{int} K)$ is a classical solution of (4.2.5). It is not difficult to check that a classical solution solves (4.2.5) in the viscosity sense: the needed property that F is increasing in X with respect to the partial ordering holds in our case.

Remark on a mnemonic rule. The monotonicity allows us to memorize easily the signs of the inequalities for F. In the smooth case for the second-order Taylor approximation, i.e., for the quadratic function $(v'(x), v''(x))$, we

[2] The reader may notice that the introduced concepts are related only with the operator and, therefore, could be called viscosity super-, sub-, and median functions, which seems to be a more natural terminology. We have no courage to deviate from the tradition already established in the theory of viscosity solutions.

have the equality. Thus, if $X \geq v''(x)$ for the pair $(v'(x), X)$ which is an element of $J^+v(x)$, we have obviously the inequality ≥ 0. Note that in the literature, the equation is quite often written with the opposite sign, and so its left-hand side is decreasing in X...

For the sake of simplicity and having in mind the specific case we shall work on, we incorporated in the definitions the requirement that the viscosity super- and subsolutions are continuous on K including the boundary. For other cases, this might be too restrictive, and more general and flexible formulations can be used.

The next criterion gives a flexibility to manipulate with the above concepts. It allows us to use smooth local majorants/minorants of a function, which is the supposed viscosity solution, as test functions (to be inserted with their derivatives into the operator).

Lemma 4.2.2 *Let $v \in C(K)$. Then the following conditions are equivalent:*

(a) *the function v is a viscosity supersolution of (4.2.5);*
(b) *for any ball $\mathcal{O}_r(x) \subseteq K$ and any $f \in C^2(\mathcal{O}_r(x))$ such that $v(x) = f(x)$ and $v \geq f$ on $\mathcal{O}_r(x)$, the inequality $\mathcal{L}f(x) \leq 0$ holds.*

Proof. (a) $\Rightarrow$ (b). Obvious: the pair $(f'(x), f''(x))$ is in $J^-v(x)$ according to the Taylor formula.

(b) $\Rightarrow$ (a). Take (p, X) in $J^-v(x)$. To conclude, we construct a smooth function f with $f'(x) = p$ and $f''(x) = X$ satisfying the requirements of (b).

By definition,

$$v(x+h) - v(x) - Q_{p,X}(h) \geq |h|^2 \varphi\big(|h|\big),$$

where $\varphi(u) \to 0$ as $u \downarrow 0$. We consider on $]0, r[$ the function

$$\delta(u) := \sup_{\{h:\ |h| \leq u\}} \frac{1}{|h|^2} \big(v(x+h) - v(x) - Q_{p,X}(h)\big)^- \leq \sup_{\{y:\ 0 \leq y \leq u\}} \varphi^-(y).$$

Obviously, δ is continuous, increasing, and $\delta(u) \to 0$ as $u \downarrow 0$. The function

$$\Delta(u) := \frac{2}{3} \int_u^{2u} \int_\eta^{2\eta} \delta(\xi)\, d\xi\, d\eta$$

vanishes at zero with its two right derivatives, and $u^2\delta(u) \leq \Delta(u) \leq u^2\delta(4u)$. It follows that the function $x \mapsto \Delta(|x|)$ belongs to $C^2(\mathcal{O}_r(0))$, its Hessian vanishes at zero, and

$$v(x+h) - v(x) - Q_{p,X}(h) \geq -|h|^2 \delta\big(|h|\big) \geq -\Delta\big(|h|\big).$$

Thus, $f(y) := v(x) + Q_{p,X}(y-x) - \Delta(|y-x|)$ is the needed function. □

For subsolutions, we have a similar result with the inverse inequalities. Using the alternative definition, we can easily establish the following:

Lemma 4.2.3 *Suppose that the function v is a viscosity solution of (4.2.5). If v is twice differentiable at x_0, then it satisfies (4.2.5) at this point in the classical sense.*

Proof. One needs to be more precise with definitions since it is not assumed that v' is defined at every point of a neighborhood of x_0. "Twice differentiable" means here that the Taylor formula at x_0 holds:

$$v(x) = v(x_0) + \langle v'(x_0), x - x_0 \rangle + \frac{1}{2}\langle v''(x_0)(x - x_0), x - x_0 \rangle + o\big(|x - x_0|^2\big).$$

Let us consider the C^2-function

$$f_\varepsilon(x) = v(x_0) + \langle v'(x_0), x - x_0 \rangle + \frac{1}{2}\langle v''(x_0)(x - x_0), x - x_0 \rangle + \varepsilon|x - x_0|^2$$

with $f_\varepsilon(x_0) = v(x_0)$. If $\varepsilon < 0$, then $v \geq f_\varepsilon$ in a sufficiently small neighborhood of x_0. Thus, by virtue of the previous lemma $\mathcal{L}f_\varepsilon(x_0) \leq 0$. Letting ε tend to zero, we obtain that $\mathcal{L}v(x_0) \leq 0$. Taking in the above definition $\varepsilon > 0$, we get the opposite inequality. □

Obviously, one can give a slightly different formulation saying that v is a viscosity supersolution of the second-order differential equation if and only if, for every $x \in \operatorname{int} K$, the inequality

$$F\big(\phi''(x), \phi'(x), v(x), x\big) \leq 0 \tag{4.2.7}$$

holds for any C^2-function ϕ such that, at the point x, the difference $v - \phi$ attains its local minimum equal to zero. The reader may ask why we replace in the inequality $\phi(x)$ by $v(x)$, which is the same number. This has sense! We can skip in the suggested reformulation the words "equal to zero" due to the following assertion, which happens to be useful in the sequel.

Lemma 4.2.4 *A function $v \in C(K)$ is a viscosity supersolution of (4.2.5) if and only if, for every point $x \in \operatorname{int} K$, inequality (4.2.7) holds for any C^2-function ϕ defined in a neighborhood of the point x and such that the difference $v - \phi$ attains its local minimum at x.*

Proof. In one direction the claim is trivial, and we need to check only that, for a supersolution, the mentioned inequality (4.2.7) holds when $v - \phi$ has a local minimum at x, i.e., when, for all y from a certain neighborhood $\mathcal{O}_\varepsilon(x)$, we have the bound

$$v(y) - \phi(y) > v(x) - \phi(x), \quad y \neq x.$$

Let $\bar{v}$ be a C^2-function dominated by v, and let g be a smooth function on $\mathbf{R}_+$ taking values in the interval $[0, 1]$ and such that $g(t) = 1$ for $t \leq \varepsilon/2$ and

$g(t) = 0$ for $t \geq \varepsilon$. Let us consider the C^2-function $\tilde{\phi} = \tilde{\phi}(y)$ with

$$\tilde{\phi}(y) = \bigl[\phi(y) + v(x) - \phi(x)\bigr] g\bigl(|x-y|\bigr) + \bigl(1 - g\bigl(|x-y|\bigr)\bigr)\bar{v}(y).$$

The difference $v - \tilde{\phi}$ attains its minimal value equal to zero at point x, and, therefore, by the supersolution property, (4.2.7) holds for $\tilde{\phi}$ and, hence, for ϕ because the two derivatives of both functions coincide at x. □

Again, a corresponding result holds for subsolutions. Notice also that specific features of the set K (with nonempty interior) were not used in the above discussions.

Now we give an application of the last characterization of the viscosity solution to prove an assertion claiming that, for a "regular" ordinary differential equation, a C^1-function known to be the viscosity solution is, in fact, a smooth one satisfying the equation in the classical sense. In the present context, the "regular" means, roughly speaking, that the equation can be solved with respect to the second derivative and the resulting right-hand side is continuous in all variables. More precisely, we have the following:

Lemma 4.2.5 *Let $\psi \in C^1(a,b)$ be a viscosity solution of the equation*

$$\psi''(z) = G\bigl(\psi'(z), \psi(z), z\bigr).$$

Suppose that the right-hand side here is a continuous function. Then the function $\psi \in C^2(a,b)$, and the equation holds in the classical sense.

Proof. Take a subinterval $[z_1, z_2]$ of $]a,b[$ and consider on it the C^2-function $\psi_\varepsilon(z)$ such that

$$\psi''_\varepsilon(z) = G\bigl(\psi'(z), \psi(z), z\bigr) + \varepsilon, \qquad \psi_\varepsilon(z_i) = \psi(z_i), \quad i = 1, 2.$$

Of course, this function could be expressed by an explicit formula, but we need not it. The parameter ε here is an arbitrary real number. We first argue with $\varepsilon > 0$. Suppose that $\psi - \psi_\varepsilon$ attains a local minimum at an interior point z of $[z_1, z_2]$. Then, necessarily, $\psi'_\varepsilon(z) = \psi'(z)$. According to the above criterion for the supersolution,

$$\psi''_\varepsilon(z) \leq G\bigl(\psi'_\varepsilon(z), \psi(z), z\bigr) = G\bigl(\psi'(z), \psi(z), z\bigr),$$

in contradiction with the definition of ψ_ε. Thus, the difference $\psi - \psi_\varepsilon$ is minimal at the extremities of $[z_1, z_2]$, where it is equal to zero. This means that $\psi(z) \geq \psi_\varepsilon(z)$ for all $z \in [z_1, z_2]$. Letting $\varepsilon \downarrow 0$ and noting that $\psi_\varepsilon(z) \to \psi_0(z)$ (even uniformly), we obtain the inequality $\psi(z) \geq \psi_0(z)$. Arguing in the same way with $\varepsilon < 0$ and using the subsolution property, we obtain the reverse inequality. So, $\psi = \psi_0$ on $[z_1, z_2]$. This means that ψ_0 is a classical solution on this interval, and it coincides with ψ. It is easily seen that such a property implies the claim of the lemma. □

4.2.5 Ishii's Lemma

The only result we need from the theory of viscosity solutions (or, better to say, from convex analysis) is the following simplified version of Ishii's lemma, see Crandall et al. [42] or Fleming and Soner [72].

Lemma 4.2.6 *Let v and $\tilde{v}$ be two continuous functions on an open subset $\mathcal{O} \subseteq \mathbf{R}^d$. Consider the function $\Delta(x,y) := v(x) - \tilde{v}(y) - \frac{1}{2}n|x-y|^2$ with $n > 0$. Suppose that Δ attains a local maximum at $(\widehat{x}, \widehat{y})$. Then there are symmetric matrices X and Y such that*

$$\big(n(\widehat{x}-\widehat{y}), X\big) \in \bar{J}^+ v(\widehat{x}), \qquad \big(n(\widehat{x}-\widehat{y}), Y\big) \in \bar{J}^- \tilde{v}(\widehat{y}),$$

and

$$\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 3n \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}. \tag{4.2.8}$$

In this statement, I is the identity matrix, and $\bar{J}^+ v(x)$ and $\bar{J}^- v(x)$ are values of the set-valued mappings whose graphs are closures of graphs of the set-valued mappings $J^+ v$ and $J^- v$, respectively.

Of course, if v is smooth, the claim follows directly from the necessary conditions of a local maximum (with $X = v''(\widehat{x})$, $Y = \tilde{v}''(\widehat{y})$ and the constant 1 instead of 3 in inequality (4.2.8)).

The following assertion is an easy exercise from linear algebra.

Lemma 4.2.7 *The inequality (4.2.8) implies that, for any $d \times m$ matrices B and C,*

$$\operatorname{tr}(BB'X - CC'Y) \leq 3n|B-C|^2. \tag{4.2.9}$$

Proof. For a symmetric matrix $S \geq O$ and any matrix G of appropriate dimension, $\operatorname{tr} GG'S = \operatorname{tr} G'S^{1/2}S^{1/2}G \geq 0$. Manipulating with block matrices and using this observation, we have

$$\begin{aligned} \operatorname{tr}(BB'X - CC'Y) &= \operatorname{tr}\begin{pmatrix} BB' & BC' \\ CB' & CC' \end{pmatrix}\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \\ &\leq 3n \operatorname{tr}\begin{pmatrix} BB' & BC' \\ CB' & CC' \end{pmatrix}\begin{pmatrix} I & -I \\ -I & I \end{pmatrix} \\ &= 3n \operatorname{tr}(BB' - BC' - CB' + CC') \\ &= 3n \operatorname{tr}(B-C)(B-C)' = 3n|B-C|^2, \end{aligned}$$

and the result is proven. □

Notice that $A(x) = \operatorname{diag} x A \operatorname{diag} x$. We denote by $\operatorname{diag} x$ the diagonal matrix whose entries on the diagonal are the coordinates of the vector x. Applying the above lemma with the matrices $B = \operatorname{diag} x A^{1/2}$ and $C = \operatorname{diag} y A^{1/2}$, we obtain the following inequality which we need in the sequel:

$$\operatorname{tr}\big(A(x)X - A(y)Y\big) \le 3n\big|A^{1/2}\big|^2|x-y|^2. \tag{4.2.10}$$

Remark. We can obtain the similar inequality

$$\operatorname{tr}\big(A(x)X - A(y)Y\big) \le 3n\operatorname{tr} A|x-y|^2$$

by "probabilistic" considerations using the above lemma in its simpler version with $m = 1$. Indeed, let ξ be a standard Gaussian vector-column, and let $\eta = A^{1/2}\xi$.

Applying the lemma with $B = \operatorname{diag} x\,\eta$ and $C = \operatorname{diag} y\,\eta$, we get the inequality

$$\operatorname{tr}(BB'X - CC'Y) \le 3n\big|\operatorname{diag}(x-y)\big|^2|\eta|^2.$$

It remains to take the expectation and note that $EBB' = A(x)$, $ECC' = A(y)$, and $E|\eta|^2 = \operatorname{tr} A$.

4.3 Uniqueness of the Solution and Lyapunov Functions

4.3.1 Uniqueness Theorem

The following concept plays a crucial role in the proof of a purely analytic result on the uniqueness of the viscosity solution, which we establish by a classical method of doubling variables using the Ishii lemma.

Definition. We say that a positive function $\ell \in C(K) \cap C^2(\operatorname{int} K)$ is the *Lyapunov function* if the following properties are satisfied:

(1) $\ell'(x) \in \operatorname{int} K^*$ and $\mathcal{L}_0\ell(x) \le 0$ for all $x \in \operatorname{int} K$,
(2) $\ell(x) \to \infty$ as $|x| \to \infty$.

Theorem 4.3.1 *Suppose that there exists a Lyapunov function ℓ. Then the Dirichlet problem (4.2.5)–(4.2.6) has at most one viscosity solution in the class of continuous functions satisfying the growth condition*

$$W(x)/\ell(x) \to 0, \quad |x| \to \infty. \tag{4.3.1}$$

Proof. Let W and $\tilde W$ be two viscosity solutions of (4.2.5) coinciding on the boundary ∂K. Suppose that $W(z) > \tilde W(z)$ for some $z \in K$. Take $\varepsilon > 0$ such that

$$W(z) - \tilde W(z) - 2\varepsilon\ell(z) > 0.$$

We introduce the family of continuous functions $\Delta_n : K \times K \to \mathbf{R}$ by putting

$$\Delta_n(x,y) := W(x) - \tilde W(y) - \frac{1}{2}n|x-y|^2 - \varepsilon\big[\ell(x) + \ell(y)\big], \quad n \ge 0.$$

Note that $\Delta_n(x,x) = \Delta_0(x,x)$ for all $x \in K$ and $\Delta_0(x,x) \le 0$ for $x \in \partial K$. From the assumption that the function l has a higher growth rate than W we deduce that $\Delta_n(x,y) \to -\infty$ as $|x|+|y| \to \infty$. It follows that the level sets $\{\Delta_n \ge a\}$ are compacts and the function Δ_n attains its maximum. That is, there exist $(x_n,y_n) \in K \times K$ such that

$$\Delta_n(x_n,y_n) = \bar\Delta_n := \sup_{(x,y)\in K\times K} \Delta_n(x,y) \ge \bar\Delta := \sup_{x\in K} \Delta_0(x,x) > 0.$$

All (x_n,y_n) belong to the compact set $\{(x,y) : \ \Delta_0(x,y) \ge 0\}$. It follows that the sequence $n|x_n - y_n|^2$ is bounded. We continue to argue (without introducing new notation) with a subsequence along which (x_n,y_n) converge to some limit $(\widehat x, \widehat x)$. Necessarily, $n|x_n - y_n|^2 \to 0$ (otherwise we would have $\Delta_0(\widehat x,\widehat x) > \bar\Delta$). It is easily seen that $\bar\Delta_n \to \Delta_0(\widehat x,\widehat x) = \bar\Delta$. Thus, $\widehat x$ is an interior point of K, and so are x_n and y_n for sufficiently large n.

By virtue of the Ishii lemma applied to the functions $v := W - \varepsilon\ell$ and $\tilde v := \tilde W + \varepsilon\ell$ at the point (x_n,y_n), there exist matrices $X^n = (X^n_{ij})$ and $Y^n = (Y^n_{ij})$ satisfying (4.2.8) and such that

$$\big(n(x_n-y_n), X^n\big) \in \bar J^+ v(x_n), \qquad \big(n(x_n-y_n), Y^n\big) \in \bar J^- \tilde v(y_n).$$

Using the notation $p_n := n(x_n-y_n) + \varepsilon\ell'(x_n)$, $q_n := n(x_n-y_n) - \varepsilon\ell'(y_n)$, $X_n := X^n + \varepsilon\ell''(x_n)$, $Y_n := Y^n - \varepsilon\ell''(y_n)$, we may rewrite the last relations in the following equivalent form:

$$(p_n, X_n) \in \bar J^+ W(x_n), \qquad (q_n, Y_n) \in \bar J^- \tilde W(y_n). \tag{4.3.2}$$

Since W and $\tilde W$ are viscosity sub- and supersolutions,

$$F\big(X_n, p_n, W(x_n), x_n\big) \ge 0 \ge F\big(Y_n, q_n, \tilde W(y_n), y_n\big).$$

The second inequality implies that $mq_n \le 0$ for each $m \in G = (-K)\cap\partial\mathcal{O}_1(0)$. But for the Lyapunov function, $\ell'(x) \in \operatorname{int} K^*$ for $x \in \operatorname{int} K$, and, therefore,

$$mp_n = mq_n + \varepsilon m\big(\ell'(x_n) + \ell'(y_n)\big) < 0.$$

Since G is a compact, $\Sigma_G(p_n) < 0$. It follows that

$$F_0\big(X_n, p_n, W(x_n), x_n\big) + U^*(p_n) \ge 0 \ge F_0\big(Y_n, q_n, \tilde W(y_n), y_n\big) + U^*(q_n).$$

Recall that U^* is decreasing with respect to the partial ordering generated by $\mathcal{C}^*$ and, hence, also by K^*. Thus, $U^*(p_n) \le U^*(q_n)$, and we obtain the inequality

$$b_n := F_0\big(X_n, p_n, W(x_n), x_n\big) - F_0\big(Y_n, q_n, \tilde W(y_n), y_n\big) \ge 0.$$

Clearly,

$$b_n = \frac{1}{2}\sum_{i,j=1}^{d}\big(a^{ij}x_n^i x_n^j X_{ij}^n - a^{ij}y_n^i y_n^j Y_{ij}^n\big) + n\sum_{i=1}^{d}\mu^i\big(x_n^i - y_n^i\big)^2$$
$$- \frac{1}{2}\beta n|x_n - y_n|^2 - \beta\Delta_n(x_n, y_n) + \varepsilon\big(\mathcal{L}_0\ell(x_n) + \mathcal{L}_0\ell(y_n)\big).$$

By virtue of (4.2.10), the first sum is dominated by const$\times n|x_n - y_n|^2$; a similar bound for the second sum is obvious; the last term is negative according to the definition of a Lyapunov function. It follows that $\limsup b_n \le -\beta\bar{\Delta} < 0$, and we get a contradiction arising from the assumption $W(z) > \tilde{W}(z)$. □

An inspection of the arguments shows that they lead to the following slightly more general and useful comparison result.

Theorem 4.3.2 *Assume that there exists a Lyapunov function ℓ. Let W and $\tilde{W}$ be, respectively, viscosity sub- and supersolution of the equation in an open set $\mathcal{O} \subseteq K$ coinciding on $\partial\mathcal{O}$ and such that*

$$W(x) = o\big(\ell(x)\big), \quad \tilde{W}(x) = o\big(\ell(x)\big), \quad |x| \to \infty.$$

Then $W(x) \le \tilde{W}(x)$ for all $x \in \bar{\mathcal{O}}$.

Remark. The definition of a Lyapunov function does not depend on U (it is a property of the operator with $U^* = 0$), and we have the uniqueness for any utility function U for which U^* is decreasing with respect to the partial ordering induced by K^*. However, to apply the uniqueness theorem, we should know that W is not growing faster than a certain Lyapunov function.

4.3.2 Existence of Lyapunov Functions and Classical Supersolutions

Results on the uniqueness of a solution to the HJB equation are all based on work with specific Lyapunov functions. The following general considerations explain how the latter can be constructed.

Let $u \in C(\mathbf{R}_+) \cap C^2(\mathbf{R}_+ \setminus \{0\})$ be an increasing strictly concave function with $u(0) = 0$ and $u(\infty) = \infty$. Introduce the function $R := -u'^2/(u''u)$. Assume that $\bar{R} := \sup_{z>0} R(z) < \infty$.

For $p \in K^* \setminus \{0\}$, we define the function $f(x) = f_p(x) := u(px)$ on K. If $y \in K$ and $x \ne 0$, then $yf'(x) = (py)u'(px) \le 0$. The inequality is strict when $p \in \operatorname{int} K^*$.

Recall that $A(x)$ is the matrix with $A^{ij}(x) = A^{ij}x^i x^j$ and the vector $\mu(x)$ has the components $\mu^i x^i$. Suppose that $\langle A(x)p, p\rangle \ne 0$. Putting $z := px$ for brevity, we obtain by obvious transformations intended to isolate full square that

$$\mathcal{L}_0 f(x) = \frac{1}{2}\Big[\big\langle A(x)p,p\big\rangle u''(z) + 2\big\langle \mu(x),p\big\rangle u'(z) + \frac{\langle \mu(x),p\rangle^2}{\langle A(x)p,p\rangle}\frac{u'^2(z)}{u''(z)}\Big] + \frac{1}{2}\frac{\langle \mu(x),p\rangle^2}{\langle A(x)p,p\rangle}R(z)u(z) - \beta u(z). \tag{4.3.3}$$

Since $u'' \le 0$, the expression in the square brackets is negative, and so is the whole right-hand side of the above formula if $\beta \ge \eta(p)\bar{R}$, where

$$\eta(p) := \frac{1}{2}\sup_{x\in K}\frac{\langle \mu(x),p\rangle^2}{\langle A(x)p,p\rangle}.$$

Of course, if $\langle A(x)p,p\rangle = 0$ we cannot argue in this way, but if in such a case also $\langle \mu(x),p\rangle = 0$, then $\mathcal{L}_0 f(x) = -\beta u(z) \le 0$ for any $\beta \ge 0$.

These simple observations lead us to the following existence result for Lyapunov functions:

Proposition 4.3.3 *Let $p \in \operatorname{int} K^*$. Suppose that $\langle \mu(x),p\rangle$ vanishes on the set $\{x \in \operatorname{int} K : \ \langle A(x)p,p\rangle = 0\}$. If $\beta \ge \eta(p)\bar{R}$, then f_p is a Lyapunov function.*

Let $\bar{\eta} := \sup_{p\in K^*}\eta(p)$. Note that $\eta(p) = \eta(p/|p|)$. Continuity considerations show that $\bar{\eta}$ is finite if $\langle A(x)p,p\rangle \ne 0$ for all $x \in K\setminus\{0\}$ and $p \in K^*\setminus\{0\}$. Obviously, if $\beta \ge \bar{\eta}\bar{R}$, then f_p is a Lyapunov function for $p \in \operatorname{int} K^*$.

The representation (4.3.3) is useful also in the search of classical supersolutions for the operator $\mathcal{L}$. Since $\mathcal{L}f = \mathcal{L}_0 f + U^*(f')$, it is natural to choose u related to U. For a particular case where $\mathcal{C} = \mathbf{R}^d_+$ and $U(c) = u(e_1 c)$, with u satisfying the postulated properties (except, maybe, unboundedness) and assuming, moreover, that the inequality

$$u^*\big(au'(z)\big) \le g(a)u(z) \tag{4.3.4}$$

holds, we get, using the homogeneity of $\mathcal{L}_0$, the following result.

Proposition 4.3.4 *Assume $\langle A(x)p,p\rangle \ne 0$ for all $x \in \operatorname{int} K$ and $p \in K^*\setminus\{0\}$. Suppose that (4.3.4) holds for all $a, z > 0$ with $g(a) = o(a)$ as $a \to \infty$. If $\beta > \bar{\eta}\bar{R}$, then there exists a_0 such that, for every $a \ge a_0$, the function af_p is a classical supersolution of (4.2.5) whatever is $p \in K^*$ with $p^1 \ne 0$. Moreover, if $p \in \operatorname{int} K^*$, then af_p is a strict supersolution on any compact subset of $\operatorname{int} K$.*

For the power utility function $u(z) = z^\gamma/\gamma$, $\gamma \in]0,1[$, we have

$$R(z) = \gamma/(1-\gamma) = \bar{R}$$

and $u^*(au'(z)) = (1-\gamma)a^{\gamma/(\gamma-1)}u(z)$. Therefore, inequality (4.3.4) holds with $g(a) = o(a)$, $a \to 0$.

If Y satisfies $\mathbf{H}_2$ with $\sigma^1 = 0$, $\mu^1 = 0$ (i.e., the first asset is the *numéraire*), and $\sigma^i \ne 0$ for $i \ne 1$, then, by the Cauchy–Schwarz inequality applied to

$\langle \mu(x), p \rangle$,

$$\eta(p) \le \frac{1}{2} \sum_{i=2}^{d} \left(\frac{\mu^i}{\sigma^i} \right)^2 .$$

The inequality

$$\beta > \frac{\gamma}{1-\gamma} \frac{1}{2} \sum_{i=2}^{d} \left(\frac{\mu^i}{\sigma^i} \right)^2 \tag{4.3.5}$$

(implying the relation $\beta > \bar{\eta}\bar{R}$) is a standing assumption in many studies on the consumption–investment problem under transaction costs, see Akian et al. [3] and Davis and Norman [47].

In particular, for the model with only one risky asset and the power utility function, by virtue of the above computations, we have, for the function $f(x) = au(px)$ given by $p \in K^*$ with $p^1 = 1$, that

$$\mathcal{L}_0 f(x) + U^*\big(f'(x)\big) = [\ldots] + \left(\frac{1}{2} \frac{\gamma}{1-\gamma} \frac{\mu^2}{\sigma^2} - \beta + (1-\gamma) a^{1/(\gamma-1)} \right) f(x),$$

where $[\ldots] \le 0$. This implies the following conclusion.

Proposition 4.3.5 *Suppose that, in the two-asset model with the power utility function, the Merton parameter*

$$\kappa_M := \frac{1}{1-\gamma} \left(\beta - \frac{1}{2} \frac{\gamma}{1-\gamma} \frac{\mu^2}{\sigma^2} \right) > 0.$$

Then the function

$$f(x) = \frac{1}{\gamma} \kappa_M^{\gamma-1} (px)^\gamma = \mathbf{m}(px)^\gamma \tag{4.3.6}$$

is a classical supersolution of the HJB equation whatever is $p \in K^*$ with $p^1 = 1$.

As we shall see in the next section, the existence of supersolutions has important implications for the Bellman function, ensuring, in particular, the finiteness of the latter.

4.4 Supersolutions and Properties of the Bellman Function

4.4.1 When is W Finite on K?

First, we present sufficient conditions ensuring that the Bellman function W of the considered maximization problem is finite.

Let Φ be the set of continuous functions $f : K \to \mathbf{R}_+$ increasing with respect to the partial ordering $\ge_K$ and such that, for every $x \in \operatorname{int} K$ and every $\pi \in \mathcal{A}_a^x$ the positive process $X^f = X^{f,x,\pi}$ given by the formula

$$X_t^f := e^{-\beta t} f(V_t) + J_t^\pi, \tag{4.4.1}$$

where $V = V^{x,\pi}$, is a supermartingale.

The set Φ of f with this property is convex and stable under the operation $\wedge$ (recall that the minimum of two supermartingales is a supermartingale). Any continuous function which is a monotone limit (increasing or decreasing) of functions from Φ also belongs to Φ.

Lemma 4.4.1 (a) *If $f \in \Phi$, then $W \le f$;*

(b) *if for any $y \in \partial K$, there exists $f \in \Phi$ such that $f(y) = 0$, then W is continuous on K.*

Proof. (a) Using the positivity of f, the supermartingale property of X^f, and, finally, the monotonicity of f, we get the following chain of inequalities leading to the required property:

$$EJ_t^\pi \le EX_t^f \le f(V_0) \le f(V_{0-}) = f(x).$$

(b) Recall that a concave function is locally Lipschitz continuous on the interior of its domain, i.e., on the interior of the set where it is finite. Hence, if Φ is not empty, then W is continuous (and even locally Lipschitz continuous) on $\operatorname{int} K$. The continuity at a point $y \in \partial K$ follows from the assumed property because $0 \le W \le f$. □

Lemma 4.4.2 *Let $f : K \to \mathbf{R}_+$ be a function in $C(K) \cap C^2(\operatorname{int} K)$. If f is a classical supersolution of (4.2.5), then $f \in \Phi$, i.e., X^f is a supermartingale.*

Proof. First, notice that a classical supersolution is increasing with respect to the partial ordering $\ge_K$. Indeed, by the finite increments formula we have that, for any $x, h \in \operatorname{int} K$,

$$f(x+h) - f(x) = f'(x + \vartheta h)h$$

for some $\vartheta \in [0,1]$. The right-hand side is greater or equal to zero because, for the supersolution f, we have the inequality $\Sigma_G(f'(y)) \le 0$ whatever is $y \in \operatorname{int} K$, or, equivalently, $f'(y)h \ge 0$ for every $h \in K$, just by the definition of the support function Σ_G and the choice of G as a generator of the cone $-K$. By continuity, $f(x+h) - f(x) \ge 0$ for every $x, h \in K$.

In order to be able to apply the Itô formula in a comfortable way, we introduce the process $\tilde V = V^{\sigma-} = VI_{[0,\sigma[} + V_{\sigma-}I_{[\sigma,\infty[}$, where σ is the first hitting time of zero by the process V. This process coincides with V on $[0,\sigma[$ but, in contrast to the latter, either always remains in $\operatorname{int} K$ (due to the stopping at σ if $V_{\sigma-} \in \operatorname{int} K$) or exits to the boundary in a continuous way and stops there. Let $\tilde X^f$ be defined by (4.4.1) with V replaced by $\tilde V$. Since

$$X^f = \tilde X^f + e^{-\beta\sigma}\big(f(V_{\sigma-} + \Delta B_\sigma) - f(V_{\sigma-})\big)I_{[\sigma,\infty[},$$

by the monotonicity of f it is sufficient to verify that $\tilde{X}^f$ is a supermartingale.

Applying Itô's formula to $e^{-\beta t} f(\tilde{V}_t)$, we obtain on $[0, \sigma[$ the representation

$$\tilde{X}_t^f = f(x) + \int_0^t e^{-\beta s} \big[\mathcal{L}_0 f(V_s) - c_s f'(V_s) + U(c_s)\big]\, ds + R_t + m_t, \quad (4.4.2)$$

where m is a process such that $m^{\sigma_n} = (m_{t\wedge\sigma_n})$ are continuous martingales for some stopping times σ_n increasing to σ, and

$$R_t := \int_0^t e^{-\beta s} f'(\tilde{V}_{s-})\, dB_s^c + \sum_{s\le t} e^{-\beta s}\big[f(\tilde{V}_{s-} + \Delta B_s) - f(\tilde{V}_{s-})\big]. \quad (4.4.3)$$

By the definition of a supersolution, for any $x \in \operatorname{int} K$,

$$\mathcal{L}_0 f(x) \le -U^*\big(f'(x)\big) \le cf'(x) - U(c) \quad \forall c \in K.$$

Thus, the integral in (4.4.2) is a decreasing process. The process R is also decreasing because the terms of the sum in (4.4.3) are less or equal to zero by monotonicity of f, while the integral is negative since

$$f'(\tilde{V}_{s-})\, dB_s^c = I_{\{\Delta B_s = 0\}} f'(\tilde{V}_{s-})\dot{B}_s\, d\|B\|_s,$$

where $f'(\tilde{V}_{s-})\dot{B}_s \le 0$ since $\dot{B}$ takes values in K. Taking into account that $\tilde{X}^f \ge 0$, we obtain from (4.4.2) that for each n, the negative decreasing process $R_{t\wedge\sigma_n}$ dominates an integrable process, and so it is integrable. The same conclusion holds for the stopped integral. Being a sum of integrable decreasing process and a martingale, the process $\tilde{X}^f_{t\wedge\sigma_n}$ is a positive supermartingale and, hence, by the Fatou lemma, $\tilde{X}^f$ is a supermartingale as well. □

Lemma 4.4.2 implies that the existence of a smooth positive supersolution f of (4.2.5) ensures the finiteness of W on K. Sometimes, e.g., in the case of power utility function, it is possible to find such a function in a rather explicit form.

Remark. Let $\bar{\mathcal{O}}$ be the closure of an open subset $\mathcal{O}$ of K, and let $f : \bar{\mathcal{O}} \to \mathbf{R}_+$ be a classical supersolution in $\bar{\mathcal{O}}$. Let $x \in \mathcal{O}$, and let τ be the exit time of the process $V^{x,\pi}$ from $\bar{\mathcal{O}}$. The above arguments imply that the process $X^f_{t\wedge\tau}$ is a supermartingale, and, therefore,

$$E\big[e^{-\beta(t\wedge\tau)} f(V_{t\wedge\tau}) + J^\pi_{t\wedge\tau}\big] \le f(x). \quad (4.4.4)$$

4.4.2 Strict Local Supersolutions

The next, slightly more technical result, the proof of which is also based on the analysis of (4.4.2), is of great importance. It will play a crucial role in deducing from the Dynamic Programming Principle that W is a subsolution of the HJB equation.

We fix a ball $\bar{\mathcal{O}}_r(x) \subseteq \operatorname{int} K$ and define τ^π as the exit time of $V^{\pi,x}$ from $\mathcal{O}_r(x)$, i.e.,

$$\tau^\pi := \inf\{t \geq 0 : |V_t^{\pi,x} - x| \geq r\}.$$

For simplicity, we assume that f is smooth in a neighborhood of $\bar{\mathcal{O}}_r(x)$.

Lemma 4.4.3 *Let $f \in C^2(\bar{\mathcal{O}}_r(x))$ be such that $\mathcal{L}f \leq -\varepsilon < 0$ on $\bar{\mathcal{O}}_r(x)$. Then there exist a constant $\eta > 0$ and an interval $]0, t_0]$ such that*

$$\sup_{\pi \in \mathcal{A}_a^x} E X_{t \wedge \tau^\pi}^{f,x,\pi} \leq f(x) - \eta t \quad \forall t \in]0, t_0].$$

Proof. We fix a strategy π and omit its symbol in the notation below. In what follows, only the behavior of the processes on $[0, \tau]$ does matter. Taking into account the monotonicity of f and modifying, if necessary, the strategy at the date τ by reducing the size of the jump ΔB_τ, we may assume without loss of generality that $|V_\tau - x| = r$ on the set $\{\tau < \infty\}$. As in the proof of Lemma 4.4.2, we apply the Itô formula. By assumption, for y from the ball $\bar{\mathcal{O}}_r(x)$, we have the bounds $\mathcal{L}_0 f(y) \leq -\varepsilon - U^*(y)$ and $\Sigma_G(f'(y)) \leq -\varepsilon$; the latter means that $kf'(y) \leq -\varepsilon|k|$ for $k \in -K$ (hence, $f'(\bar{\mathcal{O}}_r(x)) \subset \operatorname{int} K^*$). This implies the inequality

$$E X_{t \wedge \tau}^{f,x} \leq f(x) - e^{-\beta t} E N_t,$$

where

$$N_t := \varepsilon(t \wedge \tau) + \int_0^{t \wedge \tau} H(c_s, f'(V_s))\, ds + \varepsilon \int_0^{t \wedge \tau} |\dot{B}_s|\, d\|B\|_s$$

with $H(c, p) := U^*(p) + pc - U(c) \geq 0$. It remains to verify that EN_t dominates, on a certain interval $]0, t_0]$, a strictly increasing linear function which is independent of π.

Being the image of a closed ball under continuous mapping, the set $f'(\bar{\mathcal{O}}_r(x))$ is a compact in $\operatorname{int} K^*$. The lower bound of U^* on $f'(\bar{\mathcal{O}}_r(x))$ is finite. For any p from $f'(\bar{\mathcal{O}}_r(x))$ and $c \in \mathcal{C} \subseteq K$, we have the inequality $(c/|c|)p \geq \varepsilon$. At last, $U(c)/|c| \to 0$ as $c \to \infty$. Combining these facts, we infer that there is a constant κ ("large"; for convenience, $\kappa \geq 1$) such that

$$\inf_{p \in \bar{f}'(\bar{\mathcal{O}}_r(x))} H(c, p) \geq \kappa^{-1}|c|, \quad \forall c \in \mathcal{C},\ |c| \geq \kappa.$$

Thus, for the first integral in the definition of N_t, we have

$$\int_0^{t \wedge \tau} H(c_s, f'(V_s))\, ds \geq \kappa^{-1} \int_0^{t \wedge \tau} I_{\{|c_s| \geq \kappa\}} |c_s|\, ds.$$

Notice that the second integral dominates $\tilde{\kappa}\|B\|_{t \wedge \tau}$ for some $\tilde{\kappa} > 0$. To see this consider the absolute norm $|.|_1$ in $\mathbf{R}^d$. Then the total variation of B with respect to this norm is $\sum_i \operatorname{Var} B^i$, and

$$|\dot{B}|_1 = \sum_i |\dot{B}^i| = \sum_i \left| \frac{dB^i}{d\|B\|} \right| = \sum_i \left| \frac{dB^i}{d \operatorname{Var} B^i} \right| \frac{d \operatorname{Var} B^i}{d\|B\|} = \frac{d \sum_i \operatorname{Var} B^i}{d\|B\|}.$$

But all the norms in $\mathbf{R}^d$ are equivalent, i.e., $\tilde{\kappa}^{-1}|\cdot| \le |\cdot|_1 \le \tilde{\kappa}|\cdot|$ for some strictly positive constant $\tilde{\kappa}$, and the same inequalities relate the corresponding total-variation processes.

Summarizing, we conclude that it is sufficient to check the domination property for $E\tilde{N}_t$ with the simpler processes

$$\tilde{N}_t := t \wedge \tau + \int_0^{t\wedge\tau} I_{\{|c_s| \ge \kappa\}} |c_s|\, ds + \|B\|_{t\wedge\tau}. \tag{4.4.5}$$

The idea of the concluding reasoning is very simple: on a certain set of strictly positive probability, where one may neglect the random fluctuations, either τ is "large," or the total variation of the control is "large."

The formal arguments are as follows. Take $\delta > 1$. By the stochastic Cauchy formula the solution of the linear equation (4.2.2) can be written as

$$V_t^i = \mathcal{E}_t\big(Y^i\big)x^i + \mathcal{E}_t\big(Y^i\big) \int_0^t \mathcal{E}_s^{-1}\big(Y^i\big)\, d\big(B_s^i - C_s^i\big), \quad i = 1, \dots, d,$$

with the Girsanov exponential

$$\mathcal{E}\big(Y^i\big) := e^{Y^i - (1/2)\langle Y^i \rangle}.$$

Using only the fact that $\mathcal{E}_{0+}(Y^i) = \mathcal{E}_0(Y^i) = 1$, we get immediately from this representation that there exist a number $t_0 > 0$ and a measurable set Γ with $P(\Gamma) > 0$ on which

$$\big|V^{x,\pi} - x\big| \le r/2 + \delta\big(\|B\| + \|C\|\big) \quad \text{on } [0, t_0]$$

whatever is the control $\pi = (B, C)$. Of course, diminishing t_0, we may assume without loss of generality that $\kappa t_0 \le r/(4\delta)$. For any $t \le t_0$, we have on the set $\Gamma \cap \{\tau \le t\}$ the inequality $\|B\|_\tau + \|C\|_\tau \ge r/(2\delta)$, and, hence,

$$\tilde{N}_t \ge \|B\|_\tau + \|C\|_\tau - \int_0^\tau I_{\{|c_s| < \kappa\}} |c_s|\, ds \ge \frac{r}{2\delta} - \kappa t_0 \ge \kappa t_0 \ge t_0 \ge t.$$

On the set $\Gamma \cap \{\tau > t\}$, obviously, $\tilde{N}_t \ge t$. Thus, $E\tilde{N}_t \ge tP(\Gamma)$ on $[0, t_0]$, and the result is proven. □

4.5 Dynamic Programming Principle

The following property of the Bellman function is usually referred to as the (weak) "dynamic programming principle":

Theorem 4.5.1 *Assume that $W(x) < \infty$ for $x \in \operatorname{int} K$. Then for any finite stopping time τ,*

$$W(x) = \sup_{\pi \in \mathcal{A}_a^x} E\big(J_\tau^\pi + e^{-\beta\tau} W\big(V_{\tau-}^{x,\pi}\big)\big). \tag{4.5.1}$$

It is corollary of two more precise results given in Lemmas 4.5.2 and 4.5.3, which will be our tools to derive the HJB equation for the Bellman function (though nicely looking, the above formulation does not suit this purpose).

We work on the canonical filtered space of continuous functions equipped with the Wiener measure. The generic point $\omega = \omega_.$ of this space is a continuous function on $\mathbf{R}_+$, zero at the origin. Let $\mathcal{F}_t^\circ := \sigma\{\omega_s,\ s \le t\}$ and $\mathcal{F}_t := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$. We add the superscript P to denote σ-algebras augmented by all P-null sets from Ω. Recall that $\mathcal{F}_t^{\circ,P}$ coincides with $\mathcal{F}_t^P$ (this assertion follows easily from the predictable representation theorem).

A particular structure of Ω allows us to consider such operators as the stopping $\omega_. \mapsto \omega_.^s$, $s \ge 0$, where $\omega_.^s = \omega_{s\wedge .}$, and the translation $\omega_. \mapsto \omega_{s+.} - \omega_s$. Taking Doob's theorem into account, one can describe $\mathcal{F}_s^\circ$-measurable random variables as those of the form $g(w_.) = g(w_.^s)$ where g is a measurable function on Ω.

We define also the "concatenation" operator as the measurable mapping

$$g : \mathbf{R}_+ \times \Omega \times \Omega \to \Omega$$

with $g_t(s, \omega_., \tilde{\omega}_.) = \omega_t I_{[0,s[}(t) + (\tilde{\omega}_{t-s} + \omega_s) I_{[s,\infty[}(t)$.

Notice that

$$g_t\big(s, \omega_.^s, \omega_{.+s} - \omega_s\big) = \omega_t.$$

Thus, $\pi(\omega) = \pi(g(s, \omega_.^s, \omega_{.+s} - \omega_s))$.

Let π be a fixed strategy from $\mathcal{A}_a^x$, and let $\vartheta = \vartheta^{x,\pi}$ be a hitting time of zero for the process $V^{x,\pi}$.

We need the following general fact on conditional distributions.

Let ξ and η be two random variables taking values in Polish spaces X and Y equipped with their Borel σ-algebras $\mathcal{X}$ and $\mathcal{Y}$. Then ξ admits a regular conditional distribution given $\eta = y$, which we shall denote by $p_{\xi|\eta}(\Gamma, y)$, and

$$E\big(f(\xi,\eta)|\eta\big) = \int f(x,y) p_{\xi|\eta}(dx,y)\Big|_{y=\eta} \quad \text{(a.s.)}$$

for any measurable function $f(x,y) \ge 0$.

We shall apply the above relation to the random variables $\xi = (\omega_{.+\tau} - \omega_\tau)$ and $\eta = (\tau, \omega^\tau)$. In this case, according to the Dynkin–Hunt theorem, the conditional distribution $p_{\xi|\eta}(\Gamma, y)$ admits a version which is independent of y and coincides with the Wiener measure P.

At last, for fixed s and w^s, the shifted control $\pi(g(s, \omega_.^s, \tilde{\omega}.), s + dr)$ is admissible for the initial condition $V_{s-}^{x,\pi}(\omega)$. Here we denote by $\tilde{\omega}.$ a generic point of the canonical space.

Lemma 4.5.2 *Let $\mathcal{T}_f$ and $\mathcal{T}_b$ be, respectively, the sets of all finite and bounded stopping times. Then*

$$W(x) \le \sup_{\pi \in \mathcal{A}_a^x} \inf_{\tau \in \mathcal{T}_f} E\big(J_\tau^\pi + e^{-\beta\tau} W\big(V_{\tau-}^{x,\pi}\big)\big). \tag{4.5.2}$$

If $W(x) < \infty$ for all $x \in \operatorname{int} K$, then

$$W(x) \le \sup_{\pi \in \mathcal{A}_a^x} \inf_{\tau \in \mathcal{T}_b} E\big(J_\tau^\pi + e^{-\beta\tau} W\big(V_{\tau}^{x,\pi}\big)\big). \tag{4.5.3}$$

Proof. For arbitrary $\pi \in \mathcal{A}_a^x$ and $\mathcal{T}_f$, we have that

$$EJ_\infty^\pi = EJ_\tau^\pi + Ee^{-\beta\tau} \int_0^\infty e^{-\beta r} u(c_{r+\tau})\, dr.$$

According to the above discussion, we can rewrite the second term of the right-hand side as

$$Ee^{-\beta\tau} \int \left(\int_0^\infty e^{-\beta r} u\big(c_{r+\tau}\big(g\big(\tau, \omega^\tau, \tilde{\omega}\big)\big)\big)\, dr \right) P(d\tilde{\omega})$$

and dominate it by $Ee^{-\beta\tau} W(V_{\tau-}^{x,\pi})$. Thus,

$$EJ_\infty^\pi \le EJ_\tau^\pi + Ee^{-\beta\tau} W\big(V_{\tau-}^{x,\pi}\big).$$

This bound leads directly to the first announced inequality. To obtain the second, we note that W is dominated by a linear function and consider, for a bounded stopping time τ, the sequence $\tau_n := \tau + 1/n$; for τ_n, the above bound holds. Clearly, $V_{\tau_n-}^{x,\pi} \to V_\tau^{x,\pi}$. Since W is continuous in $\operatorname{int} K$ and zero is an adsorbing point, $W(V_{\tau_n-}^{x,\pi}) \to W(V_\tau^{x,\pi})$. At last, Proposition 4.2.1 allows us to apply the dominated convergence theorem and remove the annoying minus in the bound, which leads, in this modified form, to (4.5.3). □

The proof of the opposite inequality is based on different ideas.

Lemma 4.5.3 *Assume that $W(x) < \infty$ for all $x \in \operatorname{int} K$. Then for any finite stopping time τ,*

$$W(x) \ge \sup_{\pi \in \mathcal{A}_a^x} E\big(J_\tau^\pi + e^{-\beta\tau} W\big(V_{\tau-}^{x,\pi}\big)\big). \tag{4.5.4}$$

Proof. Fix $\varepsilon > 0$. Being concave, the function W is continuous on $\operatorname{int} K$. For each $x \in \operatorname{int} K$, we can find an open ball $\mathcal{O}_r(x) = x + \mathcal{O}_r(0)$ with $r = r(\varepsilon, x) < \varepsilon$ contained in the open set $\{y \in \operatorname{int} K : \ |W(y) - W(x)| < \varepsilon\}$. Moreover, we can find a smaller ball $\mathcal{O}_{\tilde{r}}(x)$ contained in the set $y(x) + K$ with $y(x) \in \mathcal{O}_r(x)$. Indeed, take a ball $x_0 + \mathcal{O}_\delta(0) \subseteq K$. Since K is a cone,

$$x + \mathcal{O}_{\lambda\delta}(0) \subseteq x - \lambda x_0 + K$$

for every $\lambda > 0$. Clearly, the requirement is met for $y(x) = x - \lambda x_0$ and $\tilde{r} = \lambda\delta$ when $\lambda|x_0| < \varepsilon$ and $\lambda\delta < r$. The family of sets $\mathcal{O}_{\tilde{r}(x)}(x)$, $x \in \operatorname{int} K$, is an open covering of $\operatorname{int} K$. But any open covering of a separable metric space contains a countable subcovering (this is the Lindelöf property; in our case, where $\operatorname{int} K$ is a countable union of compacts, it is obvious). Take a countable subcovering indexed by points x_n. For simplicity, we shall denote its elements by $\mathcal{O}_n$ and $y(x_n)$ by y_n. Put $A_1 := \mathcal{O}_1$ and $A_n = \mathcal{O}_n \setminus \bigcap_{k<n} \mathcal{O}_k$. The sets A_n are disjoint, and their union is $\operatorname{int} K$.

Let $\pi^n = (B^n, C^n) \in \mathcal{A}_a^{y_n}$ be an ε-optimal strategy for the initial point y_n, i.e., such that

$$EJ^{\pi_n} \geq W(y_n) - \varepsilon.$$

Let $\pi \in \mathcal{A}_a^x$ be an arbitrary strategy. We consider the strategy $\tilde{\pi} \in \mathcal{A}_a^x$ defined by the relation

$$\tilde{\pi} = \pi I_{[0,\tau[} + \sum_{n=1}^{\infty} \left[\left(y_n - V_{\tau-}^{x,\pi}, 0 \right) + \bar{\pi}^n \right] I_{[\tau,\infty[} I_{A_n}\left(V_{\tau-}^{x,\pi} \right) I_{\{\tau<\vartheta\}},$$

where $\bar{\pi}^n$ is the translation of the strategy π^n: namely, for a point $\omega_.$ with $\tau(\omega) = s < \infty$, we have

$$\bar{\pi}_t^n(\omega_.) := \pi_{t-s}^n(\omega_{.+s} - \omega_s).$$

In other words, the measure $d\tilde{\pi}$ coincides with $d\pi$ on $[0, \tau[$ and with the shift of $d\pi^n$ on $]\tau, \infty[$ when $V_{\tau-}^{x,\pi}$ is a subset of A_n; the correction term guarantees that in the latter case the trajectory of the control system corresponding to the control $\tilde{\pi}$ passes at time τ through the point y_n.

Now, using the same considerations as in the previous lemma, we have

$$\begin{aligned} W(x) \geq EJ_\infty^{\tilde{\pi}} &= EJ_\tau^\pi + \sum_{n=1}^{\infty} EI_{A_n}\left(V_{\tau-}^{x,\pi}\right) I_{\{\tau<\vartheta\}} \int_\tau^\infty e^{-\beta s} u\left(\bar{c}_s^n\right) ds \\ &\geq EJ_\tau^\pi + \sum_{n=1}^{\infty} EI_{A_n}\left(V_{\tau-}^{x,\pi}\right) I_{\{\tau<\vartheta\}} e^{-\beta\tau} \left(W(y_n) - \varepsilon \right) \\ &\geq EJ_\tau^\pi + Ee^{-\beta\tau} W\left(V_{\tau-}^{x,\pi}\right) - 2\varepsilon. \end{aligned}$$

Since π and ε are arbitrary, the result follows. □

Remark. The previous lemmas imply the identity

$$W(x) = \sup_{\pi \in \mathcal{A}_a^x} \inf_{\tau \in \mathcal{T}_f} E\left(J_\tau^\pi + e^{-\beta\tau} W\left(V_{\tau-}^{x,\pi} \right) \right).$$

It can be considered as another form of the dynamic programming principle.

4.6 The Bellman Function and the HJB Equation

Theorem 4.6.1 *Assume that the Bellman function W is in $C(K)$. Then W is a viscosity solution of (4.2.5).*

Proof. The claim follows from the two lemmas below. □

Lemma 4.6.2 *If (4.5.4) holds, then W is a viscosity supersolution of (4.2.5).*

Proof. Let $x \in \mathcal{O} \subseteq \operatorname{int} K$. We choose a test function $\phi \in C^2(\mathcal{O})$ such that $\phi(x) = W(x)$ and $W \geq \phi$ in $\mathcal{O}$.

At first, we fix $m \in K$ and argue with $\varepsilon > 0$ small enough to ensure that $x - \varepsilon m \in \mathcal{O}$. The function W is increasing with respect to the partial ordering generated by K. Thus,

$$\phi(x) = W(x) \geq W(x - \varepsilon m) \geq \phi(x - \varepsilon m).$$

It follows that $-m\phi'(x) \leq 0$, and, therefore, $\Sigma_G(\phi'(x)) \leq 0$.

Take now π with $B_t = 0$ and $c_t = c \in \mathcal{C}$. Let τ_r be the exit time of the continuous process $V = V^{x,\pi}$ from the ball $\bar{\mathcal{O}}_r(x) \subseteq \operatorname{int} K$. The identity (4.5.4) implies that

$$W(x) \geq E\big(J^\pi_{t\wedge\tau_r} + e^{-\beta(t\wedge\tau_r)} W(V_{t\wedge\tau_r})\big),$$

and this inequality holds true if replace W by ϕ. Writing all terms of the latter in the right-hand side and applying the Itô formula (4.4.2), we get that

$$\begin{aligned}
0 &\geq E\bigg(\int_0^{t\wedge\tau_r} e^{-\beta s} U(c_s)\, ds + e^{-\beta(t\wedge\tau_r)}\phi(V_{t\wedge\tau_r})\bigg) - \phi(x) \\
&\geq E\int_0^{t\wedge\tau_r} e^{-\beta s}\big[\mathcal{L}_0\phi(V_s) - c\phi'(V_s) + U(c)\big]\, ds \\
&\geq \min_{y\in\bar{\mathcal{O}}_r(x)} \big[\mathcal{L}_0\phi(y) - c\phi'(y) + U(c)\big] E\bigg[\frac{1}{\beta}\big(1 - e^{-\beta(t\wedge\tau_r)}\big)\bigg].
\end{aligned}$$

Dividing the resulting inequality by t and taking successively the limits as t and r converge to zero, we infer that $\mathcal{L}_0\phi(x) - c\phi'(x) + U(c) \leq 0$. Maximizing over $c \in \mathcal{C}$ yields the bound $\mathcal{L}_0\phi(x) + U^*(\phi'(x)) \leq 0$, and, therefore, W is a supersolution of the HJB equation. □

Lemma 4.6.3 *If (4.5.2) holds, then W is a viscosity subsolution of (4.2.5).*

Proof. Let $x \in \mathcal{O} \subseteq \operatorname{int} K$. Let $\phi \in C^2(\mathcal{O})$ be a function such that $\phi(x) = W(x)$ and $W \leq \phi$ on $\mathcal{O}$. Assume that the subsolution inequality for ϕ fails at x. Thus, there exists $\varepsilon > 0$ such that $\mathcal{L}\phi \leq -\varepsilon$ on some ball $\bar{\mathcal{O}}_r(x) \subseteq \mathcal{O}$. By virtue of Lemma 4.4.3 (applied to the function ϕ), there are $t_0 > 0$ and $\eta > 0$ such that on the interval $]0, t_0]$, for any strategy $\pi \in \mathcal{A}^x_a$,

$$E\big(J^\pi_{t\wedge\tau^\pi} + e^{-\beta\tau^\pi}\phi\big(V^{x,\pi}_{t\wedge\tau^\pi}\big)\big) \leq \phi(x) - \eta t,$$

where τ^π is the exit time of the process $V^{x,\pi}$ from the ball $\bar{\mathcal{O}}_r(x)$. Fix $t \in]0, t_0]$. By the second claim of Lemma 4.5.2), there exists $\pi \in \mathcal{A}_a^x$ such that

$$W(x) \le E\big(J^\pi_{t\wedge\tau} + e^{-\beta\tau} W\big(V^{x,\pi}_{t\wedge\tau}\big)\big) + \frac{1}{2}\eta t$$

for every stopping time τ, in particular, for τ^π.

Using the inequality $W \le \phi$ and applying Lemma 4.4.3, we obtain from the above relations that $W(x) \le \phi(x) - (1/2)\eta t$. This is a contradiction because at the point x the values of W and ϕ are the same. □

4.7 Properties of the Bellman Function

4.7.1 The Subdifferential: Generalities

The subdifferential of the function W at a point $x \in \operatorname{int} K$ is defined as the set

$$\partial W(x) := \big\{w \in \mathbf{R}^d : \ W(y) \le W(x) + w(y - x) \ \forall y \in K\big\}.$$

Since W is concave, this set is nonempty; obviously, it is closed and bounded. If W is unbounded, zero does not belong to $\partial W(x)$.

Recall that, for a concave function f of scalar argument, the subdifferential $\partial f(x) = [D^+ f(x), D^- f(x)]$, the interval between the values of the right and left derivatives at x.

Lemma 4.7.1 *Let x_1, x_2 be two points in* $\operatorname{int} K$. *Then*

$$\big(\partial W(x_1) - \partial W(x_2)\big)(x_1 - x_2) \le 0. \tag{4.7.1}$$

Proof. Let $w_i \in \partial W(x_i)$, $i = 1, 2$. From the definition we have the inequalities

$$W(x_2) \le W(x_1) + w_1(x_2 - x_1), \qquad W(x_1) \le W(x_2) + w_2(x_1 - x_2).$$

Adding them, we obtain that $(w_1 - w_2)(x_1 - x_2) \le 0$, the relation we need. □

Lemma 4.7.2 *The set $\partial W(x)$ is a singleton if and only if W is differentiable at x; in this case the unique element of $\partial W(x)$ is $W'(x)$.*

Lemma 4.7.3 *Let $\mathcal{O}$ be an open subset of K. The function W is of class $C^1(\mathcal{O})$ if $\partial W(x)$ is a singleton at any point $x \in \mathcal{O}$.*

Now we exploit some specific properties of the Bellman function.

The following lemma follows from the monotonicity of W with respect to the partal ordering induced by the cone K.

Lemma 4.7.4 *For every $x \in \operatorname{int} K$, we have the inclusion $\partial W(x) \subseteq K^*$.*

Proof. If $w \in \partial W(x)$, the linear function $\varphi(.) := W(x) + w(. - x)$ dominates $W(.)$ on K. Then, for any $y \in K$,

$$\varphi(x) = W(x) \le W(x+y) \le \varphi(x+y) = W(x) + wy.$$

Thus, $wy \ge 0$ for all $y \in K$, and the result follows. □

The Bellman function in the model with the power utility inherits the homotheticity property of the latter. Namely,

$$W(\nu x) = \nu^\gamma W(x) \quad \forall \nu > 0. \tag{4.7.2}$$

This implies a homotheticity property for the subdifferential.

Lemma 4.7.5 *If W satisfies (4.7.2), then*

$$\partial W(\nu x) = \nu^{\gamma-1} \partial W(x) \quad \forall \nu > 0. \tag{4.7.3}$$

Proof. Taking into account that K is a cone, we have

$$\begin{aligned}
\partial W(\nu x) &= \big\{ w \in \mathbf{R}^d : \ W(\nu y) \le W(\nu x) + w(\nu y - \nu x) \ \forall y \in K \big\} \\
&= \big\{ w \in \mathbf{R}^d : \ \nu^\gamma W(y) \le \nu^\gamma W(x) + w(\nu y - \nu x) \ \forall y \in K \big\} \\
&= \big\{ w \in \mathbf{R}^d : \ W(y) \le W(x) + \nu^{1-\gamma} w(y - x) \ \forall y \in K \big\}.
\end{aligned}$$

Since the right-hand side is $\nu^{\gamma-1}\partial W(x)$, we get the claim. □

Corollary 4.7.6 *If $W \ne 0$ satisfies (4.7.2), then $0 \notin \partial W(x)$.*

Proof. If $0 \in \partial W(x)$, then $0 \in \partial W(\nu x)$, $\nu > 0$. Thus, W attains its maximum at every point νx. In virtue of (4.7.2), this is possible only if $W = 0$. □

Lemma 4.7.7 *If W satisfies (4.7.2), then the projection of $\partial W(x)$ on $\operatorname{Lin} x$ is the singleton $\gamma W(x)|x|^{-2}x$. In particular, for $d = 2$, if not a singleton, the subdifferential $\partial W(x)$ is a closed interval orthogonal to x.*

Proof. Let $w \in \partial W(x)$, and let $w = \kappa x + w^\perp$ where $w^\perp x = 0$. Then, for any real $t > 0$, we have, in virtue of the definition of a subdifferential, that

$$W(tx) \le W(x) + wx(t-1) = W(x) + \kappa |x|^2 (t-1).$$

On the other hand, for the smooth scalar function $\psi(t) := W(tx) = t^\gamma W(x)$, the subdifferential $\partial\psi(1)$ is the singleton $\gamma W(x)$. Thus, $\gamma W(x) = \kappa |x|^2$. □

4.7.2 The Bellman Function of the Two-Asset Model

Now we investigate the structure of the Bellman function for the case $d = 2$ assuming its homotheticity. Let g_1, g_2 be the generators of K. First, let us consider the ray

$$r_1 := g_2 + \mathbf{R}_+ g_1 = \{x \in \mathbf{R}^2 : \ x = g_2 + tg_1, \ t \geq 0\},$$

parallel to g_1 and starting from the point g_2. Relation (4.7.2) allows us to recover the whole function W from its values on r_1, i.e., from the values of the concave increasing function $W_1(t) := W(g_2 + tg_1)$, $t > 0$, with $W(0+) = 0$. Its subdifferential $\partial W_1(t)$ is the interval $[D^+ W_1(t), D^- W_1(t)] \subseteq \mathbf{R}_+$; if $\tilde{t} > t$, then the interval $\partial W_1(\tilde{t})$ lays leftwards with respect to the interval $\partial W_1(t)$. Put $t_1 := \inf\{t \geq 0 : \ D^+ W_1(t) = 0\}$. Necessarily, $\partial W_1(t) = \{0\}$ for $t > t_1$.

Define the cone $K_1 := \text{cone}\{g_1, g_2 + t_1 g_1\}$ contained in K; by convention, $g_2 + \infty\, g_1 = g_1$, i.e., $K_1 := \text{cone}\{g_1\}$ when $t_1 = \infty$.

Notice that $g_1 \partial W(g_2 + tg_1) \subseteq \partial W_1(t)$. Indeed, if $w \in \partial W(g_2 + tg_1)$, then, for all $s > 0$,

$$W(g_2 + sg_1) \leq W(g_2 + tg_1) + wg_1(t - s),$$

i.e., $wg_1 \in \partial W_1(t)$.

In particular, $g_1 \partial W(g_2 + tg_1) = \{0\}$ for $t > t_1$. On the other hand, by Lemma 4.7.7 the projection of $\partial W(g_2 + tg_1)$ on the direction $g_2 + tg_1$ is a singleton. Thus, $\partial W(g_2 + tg_1)$ i also a singleton. Using Lemma 4.7.3 and Lemma 4.7.3, we arrive at the following conclusion: W is C^1 on $\text{int}\, K_1$, and $g_1 W' = 0$ on this set.

Changing the role of indices, we may introduce also the function W_2, the value t_2, and the cone $K_2 := \text{cone}\{g_2, g_1 + t_2 g_2\}$ degenerating to the ray $\text{cone}\{g_2\}$ when $t_2 = \infty$. Similarly, W is C^1 on K_2, and $g_2 W' = 0$ on this set.

Notice that $\text{int}\, K_1 \cap \text{int}\, K_2 = \emptyset$. Indeed, at a common point one would have the identities $g_i W'(x) = 0$, possible only if $W'(x) = 0$. This contradicts to Corollary 4.7.6. Therefore, $K_0 := \text{cone}\{g_2 + t_1 g_1, g_1 + t_2 g_2\}$ is a cone lying in between K_1 and K_2; the interiors of these three cones are disjoint.

Lemma 4.7.8 *For every $x \in \text{int}\, K_0$, we have the inclusion $\partial W(x) \subseteq \text{intK}^*$ or, equivalently, $wg_i > 0$ for all $w \in \partial W(x)$, $i = 1, 2$.*

Proof. As we just proved, $g_1 \partial W(g_2 + tg_1) \subseteq \partial W_1(t)$. But for $t < t_1$, the set $\partial W_1(t)$ lies in $]0, \infty[$. It follows that $g_1 \partial W(x) > 0$ for x belonging to the intersection of the ray $g_2 + \mathbf{R}_+ g_1$ with $\text{int}\, K_0$ and, hence, by Lemma 4.7.3, for all $x \in \text{int}\, K_0$. The arguments for the generator g_2 are similar. □

To check that $\text{int}\, K_1$ and $\text{int}\, K_2$ are nonempty as well as $\text{int}\, K_0$, we use more particular properties of the HJB equation. This will be done in the next section.

4.7.3 Lower Bounds for the Bellman Function

For the model where the functional depends only on the first asset which is the numéraire, one can get easily tractable lower bounds for the Bellman function which will be used later.

Let $l(x)$ be the *liquidation function*, i.e.,

$$l(x) := \sup\{z \in \mathbf{R}_+ :\ x - ze_1 \in K\}.$$

We consider the subset of admissible strategies with $\Delta B_0 = e_1 l(x) - x$ and $B_t = B_0$ for $t > 0$. This means that the agent liquidates his position in the risky asset entering the market and remains afterwards only with money. For a strategy π of this type,

$$J^{\pi}_{\infty} = \int_0^{\tau} e^{-\beta t} u(c_t)\, dt,$$

where τ is the instant when the process $X_t := l(x) - \int_0^t c_s\, ds$ hits zero. In particular, if the consumption is proportional to the wealth, i.e., $c_t = \kappa X_t$ with some constant $\kappa > 0$, we have the dynamics $X_t = l(x)e^{-\kappa t}$ with $\tau = \infty$ and

$$J^{\pi}_{\infty} = \int_0^{\infty} e^{-\beta t} u\big(\kappa l(x) e^{-\kappa t}\big)\, dt.$$

Thus,

$$W(x) \geq \sup_{\kappa>0} \int_0^{\infty} e^{-\beta t} u\big(\kappa l(x) e^{-\kappa t}\big)\, dt. \tag{4.7.4}$$

In the specific case of the power utility function,

$$J^{\pi}_{\infty} = \frac{\kappa^{\gamma}}{\gamma(\beta + \kappa\gamma)} l^{\gamma}(x).$$

The maximum of the right-hand side over κ is attained at $\kappa_* = \beta/(1-\gamma)$. This gives us a useful lower bound for the Bellman function, which we formulate as follows:

Lemma 4.7.9 *In the problem with the power utility function,*

$$W(x) \geq \frac{1}{\gamma}\kappa_*^{\gamma-1} l^{\gamma}(x) = \frac{1}{\gamma}\left(\frac{\beta}{1-\gamma}\right)^{\gamma-1} l^{\gamma}(x). \tag{4.7.5}$$

In particular,

$$W(e_1) \geq \frac{1}{\gamma}\kappa_*^{\gamma-1} = \frac{1}{\gamma}\left(\frac{\beta}{1-\gamma}\right)^{\gamma-1}. \tag{4.7.6}$$

This result will be used in the sequel for the two-asset model with the transaction cost coefficients $\lambda^{12} = \lambda^{21} = \lambda$. For such a case, at any point

$x = (\xi, \eta)$ which lies in the intersection of the solvency region K with the upper half-plane, the value of liquidation function is

$$l(x) = \xi + \frac{\eta}{1+\lambda} = \frac{p_2 x}{1+\lambda}$$

(the stock holding η is converted into $\eta/(1+\lambda)$ units of money). Therefore, we can write more explicitly that

$$W(x) \geq \frac{1}{\gamma}\kappa_*^{\gamma-1}\left(\xi + \frac{\eta}{1+\lambda}\right)^{\gamma}. \tag{4.7.7}$$

In particular, for $x = (1-z, z)$ with $z \in [0, 1+1/\lambda]$, we have the lower bound

$$W(1-z,z) \geq \frac{1}{\gamma}\left(\frac{\beta}{1-\gamma}\right)^{\gamma-1}\frac{1}{(1+\lambda)^{\gamma}}(1+\lambda-\lambda z)^{\gamma}. \tag{4.7.8}$$

Remark. Another lower bound for the Bellman function can be obtained by considering the strategy π which prescribes to convert immediately the portfolio into a single-asset one with holdings in a fixed risky asset and to consume proportionally to the current portfolio value (this means that shares permanently should be sold paying the transaction costs, which can be also interpreted as a consumption tax). Since the wealth in this case evolves accordingly to a stochastic linear equation, EJ_∞^π can be easily calculated.

4.8 The Davis–Norman Solution

4.8.1 Two-Asset Model: The Result

Let us consider the two-asset model with the price dynamics given by

$$\begin{aligned} dS_t^1 &= 0, \\ dS_t^2 &= S_t^2(\mu\, dt + \sigma\, dw_t), \end{aligned}$$

where w is a Wiener process, and $\sigma > 0$. That is, the first asset ("bond", "money", or "bank account") is the *numéraire*. The price of the risky asset follows a geometric Brownian motion. The portfolio values evolve as

$$\begin{aligned} dV_t^1 &= dL_t^{21} - \left(1+\lambda^{12}\right) dL_t^{12} - c_t\, dt, \\ dV_t^2 &= V_t^2(\mu\, dt + \sigma\, dw_t) + dL_t^{12} - \left(1+\lambda^{21}\right) dL_t^{21}, \end{aligned}$$

where L^{12} and L^{21} are adapted right-continuous increasing processes.

The optimization problem is of the form

$$E\int_0^\infty e^{-\beta s} u(c_s)\, ds \to \max, \tag{4.8.1}$$

where $u : \mathbf{R}_+ \to \mathbf{R}_+$ is a utility function. The maximum is taken over the set of strategies for which the value process evolves in the solvency cone K generated by the vectors $g_1 := (1+\lambda^{12}, -1)$ and $g_2 := (-1, 1+\lambda^{21})$ lying respectively in the forth and second quadrants. So, K and K^* are simply sectors; the generators of the dual cone are the vectors $p_1 := (1, 1+\lambda^{12})$ and $p_2 := (1+\lambda^{21}, 1)$ lying in $\mathbf{R}^2_+$ and orthogonal, respectively, to g_1 and g_2.

For the power utility function, the structure of the solution was found by Norman and Davis in 1990, though it was conjectured already in the pioneering paper of Magill and Constantinides of 1976. It was thoroughly analyzed using methods of viscosity solutions in the paper by Shreve and Soner of 1994. In our presentation we follow the latter.

We consider here the model with $u(c) = c^\gamma/\gamma$, $\gamma \in]0,1[$, supposing always that the Bellman function W is finite. As we already know, such a property is guaranteed if $\kappa_M > 0$, i.e.,

$$\beta > \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{\mu^2}{\sigma^2}. \tag{4.8.2}$$

Note that the above inequality ensures the finiteness of W even in the classical Merton problem without friction. In the model with friction one can find other sufficient conditions for the finiteness of the Bellman function. We discuss this issue later.

It follows from the general theory that the Bellman function W is the viscosity solution of the HJB equation with zero boundary condition. It is unique in the class of functions with growth rate $\gamma' \geq \gamma$ such that the above bound still holds with γ'.

We assume, moreover, that the instantaneous interest rate of the risky asset $\mu > 0$, and this hypothesis will be used immediately in proving (4.8.5).

The HJB equation can be written as follows:

$$\max\left\{\frac{1}{2}\sigma^2\eta^2 W_{\eta\eta} + \mu\eta W_\eta - \beta W + u^*(W_\xi), -g_1 W', -g_2 W'\right\} = 0. \tag{4.8.3}$$

Of course, at the moment, we have no information about the existence of the involved derivatives of W, and the above relation has to be understood in the viscosity sense. Since this section is a case study and our intention is to obtain an explicit solution, we abandon the standard notation: to improve the perception of the formulae, we use the notation (ξ, η) and (t, z) for generic points in $\mathbf{R}^2_+$. Moreover, for the sake of simplicity, we suppose that the transaction costs for buying and selling are the same, i.e., $\lambda^{12} = \lambda^{21} = \lambda > 0$.

The principal result is the following:

Theorem 4.8.1 *There are vectors $\tilde{g}_1, \tilde{g}_2$ such that the solvency cone K is the union of $K_1 = \text{cone}\{g_1, \tilde{g}_1\}$, $K_2 = \text{cone}\{g_2, \tilde{g}_2\}$, and $K_0 = \text{cone}\{\tilde{g}_1, \tilde{g}_2\}$, three sectors with disjoint nonempty interiors. The Bellman function W is a concave positive homogeneous function of order γ and belongs to the class*

$C^1(\operatorname{int} K)$. *On* K_i, $i = 1, 2$, *it is given by the formulae* $a_i u(p_i x)$. *On* $\operatorname{int} K_0$, *it is a classical* C^2*-solution of the equation*

$$\frac{1}{2}\sigma^2\eta^2 W_{\eta\eta} + \mu\eta W_\eta - \beta W + u^*(W_\xi) = 0, \tag{4.8.4}$$

where $u^*(p) = (1-\gamma)\gamma^{-1}p^{\gamma/(\gamma-1)}$.

4.8.2 Structure of Bellman Function

First of all, we recall that the functions $au(p_i x)$ for large a are classical supersolutions, and, hence, they dominate $W(x)$ "globally," i.e., on the whole K. We refine this result by showing that, on certain sectors with nonempty interiors and adjacent to the boundaries of the solvency region, the Bellman function W coincides with functions of this particular type. Specifically, we have:

Proposition 4.8.2 (a) *There exists* $a_1 > 0$ *such that*

$$W(x) = a_1 u(p_1 x) \quad \text{on } \operatorname{cone}\{g_1, e_1\}. \tag{4.8.5}$$

(b) *There exists* $a_2 > 0$ *such that*

$$W(x) = a_2 u(p_2 x) \quad \text{on } \operatorname{cone}\big\{g_2, 2\theta(1+\lambda)^{-1}g_2 + e_2\big\}. \tag{4.8.6}$$

Proof. (a) Take $a_1 := W(e_1)/u(p_1 e_1) = \gamma W(e_1)$. Due to the homotheticity property, this choice implies that the function $\varphi(x) := a_1 u(p_1 x)$ coincides with $W(x)$ on the whole ray $\mathbf{R}_+ e_1$. This immediately implies that $W(x) \geq \varphi(x)$ on the sector $\operatorname{cone}\{g_1, e_1\}$ because, along each ray $\xi e_1 + \mathbf{R}_+ g_1$, $\xi > 0$, the function W is increasing while φ remains constant. On the other hand, both functions are zero on the boundary ray $\mathbf{R}_+ g_1$ (which is a part of ∂K). Let us check that φ is a classical supersolution of our HJB equation on the sector $\operatorname{cone}\{g_1, e_1\}$, and, hence, we have the reverse inequality $W(x) \leq \varphi(x)$ on this set due to Lemmas 4.4.1 and 4.4.2. The only problem is to check, in the interior of the sector, the inequality $\mathcal{L}_0\varphi + u^*(\varphi_\xi) \leq 0$, which, in the detailed notation, is simply

$$\frac{1}{2}\sigma^2\eta^2\varphi_{\eta\eta} + \mu\eta\varphi_\eta - \beta\varphi + u^*(\varphi_\xi) \leq 0.$$

The first term in the left-hand side is always negative (due to the concavity). The second one is negative because, for $x = (\xi, \eta)$ in the considered set, the coordinate $\eta < 0$, the parameter $\mu > 0$ by assumption, and φ is increasing in η. At last, by virtue of the bound (4.7.6),

$$a_1 := \gamma W(e_1) \geq \kappa_*^{\gamma-1} = \left(\frac{\beta}{1-\gamma}\right)^{\gamma-1}, \tag{4.8.7}$$

and we easily get that

$$u^*(\varphi_\xi) = (1-\gamma)a_1^{\frac{1}{\gamma-1}}\varphi \le \beta\varphi. \tag{4.8.8}$$

(b) We put $N := 2\theta(1+\lambda)^{-1}$ and note that

$$\frac{1}{2}\sigma^2(1+\lambda)(\gamma-1)N + \mu = 0.$$

Take now $a_2 := W(Ng_2 + e_2)/u(p_2e_2) = \gamma W(Ng_2 + e_2)$. By the same general arguments as above, we obtain that the function $\psi(x) := a_2u(p_2x)$ coincides with $W(x)$ on the whole ray generated by the vector $Ng_2 + e_2$ and satisfies the inequality $W(x) \ge \psi(x)$ on the sector cone$\{g_2, Ng_2 + e_2\}$. To obtain the reverse inequality, it remains to verify that $\mathcal{L}_0\psi + u^*(\psi_\xi) \le 0$ at any point $x = (\xi, \eta)$ from the interior of this sector. Such a point admits a unique representation

$$x = \bar{\xi}g_2 + \bar{\eta}(Ng_2 + e_2)$$

with some reals $\bar{\xi}, \bar{\eta} > 0$. We have: $p_2x = \bar{\eta}p_2e_2 = \bar{\eta}$ and

$$\eta = xe_2 = \bar{\xi}g_2e_2 + \bar{\eta}(Ng_2 + e_2)e_2 \ge N\bar{\eta}g_2e_2 = N\bar{\eta}(1+\lambda).$$

Thus,

$$\begin{aligned}\frac{1}{2}\sigma^2\eta^2\psi_{\eta\eta}(x) + \mu\eta\psi_\eta(x) &= a_2(p_2x)^{\gamma-1}\eta\left(\frac{1}{2}\sigma^2(\gamma-1)\frac{\eta}{p_2x} + \mu\right)\\ &\le a_2(p_2x)^{\gamma-1}\eta\left(\frac{1}{2}\sigma^2(\gamma-1)N(1+\lambda) + \mu\right) = 0\end{aligned}$$

due to our choice of N.

On the other hand, the value of liquidation function is $l(Ng_2 + e_2) = 1/(1+\lambda)$ (the intersection of the ray $Ng_2 + e_2 - \mathbf{R}_+g_2$ and the axis of abscises is the point $(1/(1+\lambda), 0)$), and, therefore, due to the bound (4.7.6), we have that

$$a_2 \ge \kappa_*^{\gamma-1}\frac{1}{(1+\lambda)^\gamma} = \left(\frac{\beta}{1-\gamma}\right)^{\gamma-1}\frac{1}{(1+\lambda)^\gamma}. \tag{4.8.9}$$

It follows that

$$u^*\big(\psi_\xi(x)\big) = (1-\gamma)a_2^{\frac{1}{\gamma-1}}(1+\lambda)^{\frac{\gamma}{\gamma-1}}\psi(x) \le \beta\psi(x), \tag{4.8.10}$$

and we get the result. □

We denote by K_i, $i = 1, 2$, the largest sectors on which the Bellman function is given by the formulae $W(x) = a_i(p_ix)^\gamma$. By the above we have:

Corollary 4.8.3 *The sectors K_i, $i = 1, 2$, are nonempty, and*

$$a_1 \ge \kappa_*^{\gamma-1}, \qquad a_2 \ge \kappa_*^{\gamma-1}\frac{1}{(1+\lambda)^\gamma}. \tag{4.8.11}$$

Our next aim is to show that $\operatorname{int} K_0 \neq \emptyset$ or, in other words, that K_1 and K_2 have no common boundary points (except zero). Though the crucial information will be obtained by a reduction to a one-dimensional problem, we make the first step in this direction immediately by inspecting the above formulae and establishing the following simple assertion:

Lemma 4.8.4 *If*

$$W(e_1) = \frac{1}{\gamma}\left(\frac{\beta}{1-\gamma}\right)^{\gamma-1}, \tag{4.8.12}$$

then the axis of abscises is not the common boundary of K_1 and K_2.

Proof. Suppose the opposite. Then the function $\psi(x) = a_2(p_2x)^\gamma$ coincides with $W(x)$ on the sector $\operatorname{cone}\{g_2, e_1\}$, and we can determine the value of a_2 using (4.8.12). It corresponds to the equalities in (4.8.9) and (4.8.10). Hence, in the considered case,

$$\mathcal{L}_0\psi(x) + u^*\big(\psi_\xi(x)\big) = a_2(p_2x)^{\gamma-1}\eta\left(\frac{1}{2}\sigma^2(\gamma-1)\frac{\eta}{p_2x} + \mu\right).$$

The right-hand side is strictly positive for points $x = (\xi, \eta)$ with sufficiently small coordinate $\eta > 0$. This means that, for such points, ψ cannot be the solution of the HJB equation. □

Usually, to find the constants a_i, one has to solve a free-boundary problem. However, in the special case where $\theta = 1$, the value a_2 can be calculated easily. The optimal strategy is just to sell a constant proportion of the stock to generate a flow of money for consumption. The precise result is here.

Proposition 4.8.5 *Suppose that the Merton parameter*

$$\kappa_M := \frac{1}{1-\gamma}\left(\beta - \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{\mu^2}{\sigma^2}\right) > 0$$

and $\theta := (1-\gamma)^{-1}\mu\sigma^{-2} = 1$ (i.e., $\mu = (1-\gamma)\sigma^2$).
Then

$$W(x) = \mathbf{m}\frac{1}{(1+\lambda)^\gamma}(p_2x)^\gamma \quad \text{on } K \cap \big\{x = (\xi, \eta) :\ \xi \le 0\big\}. \tag{4.8.13}$$

Proof. Let us consider the process $V = (V^1, V^2)$ with $V^1 = 0$ and

$$dV_t^2 = V_t^2(\mu\, dt + \sigma\, dw_t) - \kappa_M V_t^2\, dt, \qquad V_0^2 = 1.$$

It corresponds to the strategy when the agent, having as the initial endowment a unit of stock, converts instantaneously a constant proportion of his wealth into cash and uses it immediately for the consumption with intensity given by the formula $c_t = (1+\lambda)^{-1}\kappa_M V_t^2$.

Apparently,

$$E\big(V_t^2\big)^\gamma = e^{\gamma(\mu-\sigma^2/2-\kappa_M)t+(1/2)\gamma^2\sigma^2 t} = e^{(\beta-\kappa_M)t},$$

the second equality holding because of the assumed identity $\mu = (1-\gamma)\sigma^2$. Thus, for the considered strategy,

$$J_\infty^\pi = \int_0^\infty e^{-\beta t} Eu(c_t)\,dt = \frac{1}{\gamma}\kappa_M^{\gamma-1}\frac{1}{(1+\lambda)^\gamma} = \mathbf{m}\frac{1}{(1+\lambda)^\gamma} = f(e_2),$$

where we denote by f the right-hand side of (4.8.13). It follows that $W(e_2) \geq f(e_2)$. In fact, we have the equality here since we know already that f is a classical supersolution (see Proposition 4.3.5), and, hence, $W \leq f$ in K. By the homotheticity, $W = f$ on $\mathbf{R}_+ e_2$. Since along each ray $\eta e_2 - \mathbf{R}_+ g_2$, $\eta > 0$, the function f is constant while W is decreasing, we have, on the sector $K \cap \{x = (\xi,\eta) :\ \xi \leq 0\}$, the inequality $W \leq f$ and, hence, the equality. □

4.8.3 Study of the Scalar Problem

The utility function is homogeneous of degree γ, and this property, due to the linearity of the dynamics, is inherited by the Bellman function, i.e.,

$$W(x) = \nu^\gamma W(x/\nu) \quad \forall \nu > 0. \tag{4.8.14}$$

Thus, knowing W on the intersection of the line $\{(\xi,\eta) : \xi + \eta = 1\}$ with the interior of K, that is, on the interval with the extremities $(-1/\lambda, 1+1/\lambda)$ and $(1+1/\lambda, -1/\lambda)$, one can reconstruct this function, using the homotheticity property, on the whole domain by the formula

$$W(\xi,y) = (\xi+\eta)^\gamma W\bigg(\frac{\xi}{\xi+\eta}, \frac{\eta}{\xi+\eta}\bigg), \quad (\xi,\eta) \in \operatorname{int} K. \tag{4.8.15}$$

Let us consider the bijection mapping $T : (\xi,\eta) \mapsto (\xi = \eta, \eta/(\xi+\eta))$ of $\operatorname{int} K$ onto the rectangular $]0,\infty[\,\times\,]-1/\lambda, 1+1/\lambda[$; clearly, $T \in C^\infty$. It follows that the function $\Phi(t,z) = t^\gamma \psi(z)$ with $\psi(z) := W(1-z,z)$ is a viscosity solution of the equation obtained by the change of variables.

Specifically, let $t = t(\xi,\eta) = \xi + \eta$, $z = z(\xi,\eta) = \eta/(\xi+\eta)$ (and, hence, $\xi = t(1-z)$, $\eta = tz$). Differentiating the identity

$$W(\xi,\eta) = \big[t(\xi,\eta)\big]^\gamma \psi\big(z(\xi,\eta)\big),$$

we obtain the following formulae for derivatives:

$$\begin{aligned} W_\xi(\xi,\eta) &= t^{\gamma-1}\big[\gamma\psi(z) - z\psi'(z)\big], \\ W_\eta(\xi,\eta) &= t^{\gamma-1}\big[\gamma\psi(z) + (1-z)\psi'(z)\big], \\ W_{\eta\eta}(\xi,\eta) &= t^{\gamma-2}\big[\gamma(\gamma-1)\psi(z) + 2(\gamma-1)(1-z)\psi'(z) + (1-z)^2\psi''(z)\big]. \end{aligned}$$

Therefore, we obtain that

$$g_1 W'(\xi,\eta) = t^{\gamma-1}\big[\lambda\gamma\psi(z) - (1+\lambda z)\psi'(z)\big],$$
$$g_2 W'(\xi,\eta) = t^{\gamma-1}\big[\lambda\gamma\psi(z) + (1+\lambda-\lambda z)\psi'(z)\big].$$

The formal substitution into (4.8.3) yields the following equation in the viscosity sense on the interval $[-1/\lambda, 1+1/\lambda]$ for the continuous function ψ vanishing at the extremities:

$$\max_{0\le i\le 2} \ell_i\psi = 0 \tag{4.8.16}$$

with two first-order operators

$$\ell_1\psi := -\lambda\gamma\psi + (1+\lambda z)\psi', \qquad \ell_2\psi := -\lambda\gamma\psi - (1+\lambda-\lambda z)\psi',$$

and the second-order operator

$$\ell_0\psi = f_2\psi'' + f_1\psi' + f_0\psi + \frac{1-\gamma}{\gamma}[\gamma\psi - z\psi']^{\frac{\gamma}{\gamma-1}},$$

where

$$f_2(z) := \frac{1}{2}\sigma^2 z^2(1-z)^2,$$
$$f_1(z) := -\sigma^2(1-\gamma)z(1-z)(z-\theta),$$
$$f_0(z) := \frac{1}{2}\sigma^2\gamma(\gamma-1)z^2 + \gamma\mu z - \beta,$$

and $\theta := (1-\gamma)^{-1}\mu\sigma^{-2}$ is the Merton proportion.

The function ψ, being concave, has left and right derivatives continuous from the left and right, respectively, and satisfying the inequality $D^+\psi \le D^-\psi$, which can be strict only on a countably set. Outside this set, the derivative ψ' exists and is continuous.

Moreover, ψ is twice differentiable (in the sense of the Taylor formula) almost everywhere, and, therefore, (4.8.16) holds in the classical sense almost everywhere, see Lemma 4.2.3. This means that, at each point outside an exceptional null-set, we have three inequalities

$$\ell_1\psi(z) \le 0, \qquad \ell_0\psi(z) \le 0, \qquad \ell_2\psi(z) \le 0, \tag{4.8.17}$$

and at least one of them is "active," that is, holds with the equality. By continuity, on the whole interval,

$$-\lambda\gamma\psi + (1+\lambda z)D^{\pm}\psi \le 0, \qquad -\lambda\gamma\psi - (1+\lambda-\lambda z)D^{\pm}\psi \le 0. \tag{4.8.18}$$

Lemma 4.8.6 *The function ψ is continuously differentiable on the interval $I :=]-1/\lambda, 1+1/\lambda[$ except, maybe, zero. If ψ' has a discontinuity at zero, then*

$$\psi(0) = \frac{1}{\gamma}\left(\frac{1-\gamma}{\beta}\right)^{1-\gamma} = \frac{1}{\gamma}\kappa_*^{1-\gamma}. \tag{4.8.19}$$

Proof. Since ψ is concave, it has left and right derivatives satisfying the inequality $D^+\psi \le D^-\psi$. Suppose that at a point z the inequality is strict. Let $p \in]D^+\psi(z), D^-\psi(z)[$. It follows that $(p, X) \in J^+\psi(z)$ whatever is $X \in \mathbf{R}$. By the definition of viscosity subsolution, we should have at least one of the following three inequalities: $\ell_i Q_{p,X}(z) \ge 0$, $i = 0, 1, 2$. This leads to an immediate contradiction if $z \ne 0$ or $z \ne 1$. Indeed, the coefficient at the second derivative being strictly positive, $\ell_0 Q_{p,X}(z) \to -\infty$ as $X \to -\infty$. A nonconstant linear function negative at the extremities of an interval is strictly negative in its interior, and, therefore, (4.8.18) implies that $\ell_i Q_{p,X}(z) < 0$ for $i = 1, 2$.

For the point $z = 1$, we can say only that

$$\ell_0 Q_{p,n}(z) = f_0(1)\psi(1) + u^*\big(\gamma\psi(1) - p\big) \ge 0. \tag{4.8.20}$$

To obtain a contradiction, we recall the classical fact that, for the monotone function ψ', the derivative ψ'' exists almost everywhere. Another fact (less known) is that ψ'' is locally integrable. On the other hand, the function $1/f_2$ has a nonintegrable singularity at 1. With this, we can find a sequence $z_n \uparrow 1$ (or $z_n \downarrow 1$) such that $\psi''(z_n)$ does exist, $\lim_n f_2(z_n)\psi''(z_n) = 0$, and inequalities (4.8.17) hold at z_n. The passage to the limit in the central one yields the inequalities

$$f_0(1)\psi(1) + u^*\big(\gamma\psi(1) - D^\pm\psi(1)\big) \le 0.$$

The function u^* being strictly monotone,

$$f_0(1)\psi(1) + u^*\big(\gamma\psi(1) - p\big) < 0,$$

contradicting to (4.8.20).

The above arguments for $z = 0$ do not work. An attempt to repeat them leads to a conclusion that if ψ' has a discontinuity at zero, then, necessarily,

$$f_0(0)\psi(0) + u^*\big(\gamma\psi(0)\big) = 0.$$

Solving this equation, we get the formula (4.8.19). □

Remark. On $\operatorname{int} K$, the Bellman function W always has a continuous derivative in the radial direction. Thus, the second claim of the lemma means that if W has no derivative in transversal directions at the ray $\mathbf{R}_+ e_1$, then, necessarily,

$$W(\xi, 0) = \frac{1}{\gamma}\left(\frac{1-\gamma}{\beta}\right)^{1-\gamma} \xi^\gamma.$$

Due to the continuity of the derivative, we may guess that the regions where inequalities in (4.8.16) are active are intervals. The concavity of ψ makes plausible a more specific structure of ψ, namely, that this function satisfies the above three differential inequalities, the first is a differential equation on the interval $]-1/\lambda, z_1[$, the second on $]z_1, z_2[$, and the third on $]z_2, 1+1/\lambda[$. The first-order differential equations $\ell_i\psi = 0$ with zero boundary conditions can be readily solved, and we have the following explicit formulae for the external intervals:

$$\psi(z) = \kappa_1(1+\lambda z)^\gamma, \quad z \in [-1/\lambda, z_1],$$
$$\psi(z) = \kappa_2(1+\lambda-\lambda z)^\gamma, \quad z \in [z_2, 1+1/\lambda],$$

where the constants $\kappa_i > 0$ are to be specified.

Lemma 4.8.7 *The interior of K_0 is nonempty.*

Proof. Suppose the opposite. This means that there exists a point x different from the origin which belongs to the common boundary of K_1 and K_2. If the function W is differentiable at x, we obtain that $W^{1/\gamma}$ has vanishing partial derivatives in the directions g_1 and g_2. It follows that $W'(x) = 0$, which is impossible. If W is not differentiable at x, then x is on the axis of abscises, and we refer simply to Lemmas 4.8.4 and 4.8.6. □

The most difficult part of the analysis is to show the following result claiming that the axis of abscises is not on the boundary of K_0.

Proposition 4.8.8 *The point e_1 belongs to* $\operatorname{int} K_1$.

Proof. Suppose that the assertion is note true, that is, the value $z_1 = 0$. If $\gamma W(e_1) > \kappa_*^{\gamma-1}$, then, as we know, $W \in C^1(\operatorname{int} K)$. In a right neighborhood of zero, $\psi(z) = W(1-z, z)$ is the solution of the second-order differential equation $\ell_0\psi = 0$. Since $f_1(0) = 0$ and $f_0(0) = -\beta$, we have that

$$\lim_{z\downarrow 0} \frac{1}{2}\sigma^2 z^2 \psi''(z) = \beta\psi(0) - \frac{1-\gamma}{\gamma}\big[\gamma\psi(0)\big]^{\frac{\gamma}{\gamma-1}}.$$

Noticing that the derivative of the function

$$H(y) := \beta y - \frac{1-\gamma}{\gamma}(\gamma y)^{\frac{\gamma}{\gamma-1}}, \quad y > 0,$$

is strictly positive and $H(\kappa_*^{\gamma-1}/\gamma) = 0$, we conclude that the limit above is also strictly positive. But this is impossible because ψ is concave and $\psi'' \le 0$.

Consider the "critical" case where $\gamma W(e_1) = \kappa_*^{\gamma-1}$ and the first derivative of ψ has a jump downwards at point zero.

Now we know the function $\psi(z)$ on the interval $[-1/\lambda, 0]$ explicitly. On the other hand, the lower bound for W implies the inequality

$$\psi(z) \ge h(z) := \frac{1}{\gamma}\kappa_*^{\gamma-1}\frac{1}{(1+\lambda)^\gamma}(1+\lambda-\lambda z)^\gamma.$$

In a right neighborhood of zero, $\psi(z)$ is the solution of the second-order differential equation. The difference $\tilde\psi(z) = \psi(z) - h(z) \ge 0$, and $\tilde\psi(0) = 0$. It is clear that

$$0 \le D^+\tilde{\psi}(0) = D^+\psi(0) - h'(0) \le D^-\psi(0) - h'(0) < \infty.$$

Substitution of $\psi(z) = h(z) + \tilde{\psi}(z)$ into the equation $\ell_0\psi(z) = 0$ yields the identity

$$g_1(z) + g_2(z) + g_3(z) = 0,$$

where

$$\begin{aligned} g_1(z) &:= \ell_0 h(z), \\ g_2(z) &:= \ell_0\tilde{\psi}(z) - u^*\big(\gamma\tilde{\psi}(z) - z\tilde{\psi}'(z)\big), \\ g_3(z) &:= u^*\big(\gamma h(z) - zh'(z) + \gamma\tilde{\psi}(z) - z\tilde{\psi}'(z)\big) - u^*\big(\gamma h(z) - zh'(z)\big). \end{aligned}$$

Note that

$$g_1(z) = f_2(z)h''(z) + f_1(z)h'(z) + \big(f_0(z) + \beta\big)h(z).$$

It follows that $g_1(0) = 0$ and

$$g_1'(0) = f_1'(0)h'(0) + f_0'(0)h(0) = \frac{\mu}{1+\lambda}\kappa_*^{\gamma-1}.$$

Observing that also $g_3(0+) = 0$, we infer that $g_2(0+) = 0$ as well, and, hence,

$$\lim_{z\downarrow 0} z^2\tilde{\psi}''(z) = 0. \tag{4.8.21}$$

The existence of the derivatives of g_1 and g_3 implies the existence of the derivatives of g_2. We have the identity for the derivatives

$$g_1'(z) + g_2'(z) + g_3'(z) = 0, \tag{4.8.22}$$

implying that

$$\lim_{z\downarrow 0}\big[g_2'(z) + g_3'(z)\big] = -\lim_{z\downarrow 0} g_1'(z) = -\frac{\mu}{1+\lambda}\kappa_*^{\gamma-1}. \tag{4.8.23}$$

The differentiability of g_2 implies that the derivative $\tilde{\psi}'''(z)$ does exist and is continuous for $z > 0$. Differentiating the expression

$$g_2(z) = f_2(z)\tilde{\psi}''(z) + f_1(z)\tilde{\psi}'(z) + f_0(z)\tilde{\psi}(z),$$

we obtain that

$$\lim_{z\downarrow 0} g_2'(z) = (\mu - \beta)D^+\tilde{\psi}(0) + \lim_{z\downarrow 0}\big[(1/2)\sigma^2 z^2\tilde{\psi}'''(z) + \big(\sigma^2 + \mu\big)z\tilde{\psi}''(z)\big].$$

Differentiating the formula

$$g_3(z) = \frac{1-\gamma}{\gamma}\left[\kappa_*^{\gamma-1}\left(\frac{1+\lambda-\lambda z}{1+\lambda}\right)^{\gamma-1} + \gamma\tilde{\psi}(z) - z\tilde{\psi}'(z)\right]^{\frac{\gamma}{\gamma-1}} - \frac{1-\gamma}{\gamma}\kappa_*^{\gamma}\left(\frac{1+\lambda-\lambda z}{1+\lambda}\right)^{\gamma},$$

we infer that

$$\lim_{z\downarrow 0} g_3'(z) = -\kappa_*(\gamma-1)D^+\tilde{\psi}(0) + \kappa_* \lim_{z\downarrow 0} z\tilde{\psi}''(z).$$

Adding this identity with that for the limit of $g_2'(z)$, we arrive at the formula

$$\lim_{z\downarrow 0}\big[g_2'(z)+g_3'(z)\big] = \mu D^+\tilde{\psi}(0) + \lim_{z\downarrow 0}\big[(1/2)\sigma^2 z^2\tilde{\psi}'''(z) + \big(\sigma^2+\mu+\kappa_*\big)z\tilde{\psi}''(z)\big].$$

In virtue of the lemma below, the right-hand side is positive, in contradiction with identity (4.8.23). □

Lemma 4.8.9 *Let f be a bounded C^2-function on the interval $]0,\varepsilon[$, and let $\kappa \in \mathbf{R}$. Then*

$$\limsup_{z\downarrow 0}\big[z^2 f''(z) + \kappa z f'(z)\big] \ge 0.$$

Proof. Put $z := e^{-t}$ and consider the bounded function $\tilde{f}(t) = f(e^{-t})$. The claimed property means that

$$\limsup_{t\to\infty}\big[\tilde{f}''(t) + \kappa\tilde{f}'(t)\big] \ge 0$$

whatever is the constant κ. Suppose that the assertion fails. Then there exists $\kappa_1 > 0$ such that

$$\tilde{f}''(t) + \kappa\tilde{f}'(t) \le -\kappa_1$$

for all sufficiently large t. The integration yields the inequality

$$\tilde{f}'(t) + \kappa\tilde{f}(t) \le \kappa_2 - \kappa_1 t,$$

leading to an obvious contradiction: a function cannot be bounded while its derivative converges to $-\infty$. □

Proposition 4.8.8 completes the proof of Theorem 4.8.1. It provides the information that the suspicious point $z = 0$ belongs to the interval where the function ψ, given by the explicit formula, is smooth. Thus, ψ is C^1.

4.8.4 Skorohod Problem

In the classical Merton problem there are no difficulties to construct the optimal pair: the optimal wealth process is a solution of a simple linear stochastic equation, the optimal control is a linear function of the solution, and the same linear equation describes the optimal dynamics. In the model with transaction costs the situation is much more complicated. The optimal pair (i.e., the portfolio process and the control) is a solution of the stochastic Skorokhod problem (called also a stochastic differential equation with reflection). Moreover, the needed particular case of this problem has rather unpleasant features: the domain is a sector (so the boundary is not smooth), reflection is oblique, and the explicit form of the drift coefficient is not available. Since differential equations with reflections are rarely treated in the monographic literature, we provide in the Appendix a brief introduction with an elementary result which well serves our purpose here. In this subsection we use this result to check the existence and uniqueness of the optimal pair in the considered optimal control problem.

We have established that the solvency cone K can be decomposed into the union of three convex cones K_i (sectors, in fact) with disjoint nonempty interiors. The sectors K_i, $i = 1, 2$, share their "external" boundaries $\mathbf{R}_+ g_i$ with the solvency cone K, while the "internal" boundaries form the boundaries of K_0. The function $W^{1/\gamma}$ is linear in K_1 and K_2. Moreover, the axis of abscises is in the interior of K_1. The Bellman function W is C^1 in $\operatorname{int} K$.

Let $g : \partial K_0 \to \mathbf{R}^2$ be a vector-valued function with $g(x) = -g_i$ on the set $(\partial K_0 \cap \partial K_i) \setminus \{0\}$ and $g(0) = 0$. We consider on K_0 the Skorokhod problem formulated as follows: find a pair of adapted continuous processes, V, starting from $x \in K_0$, evolving in K_0, and trapped at zero, and k, scalar, starting at zero, and increasing, such that

$$dV_t^1 = -\big(W_\xi(V_t)\big)^{1/(\gamma-1)}\,dt + g_1(V_t)\,dk_t, \tag{4.8.24}$$

$$dV_t^2 = V_t^1(\mu\,dt + \sigma\,dw_t) + g_2(V_t)\,dk_t, \tag{4.8.25}$$

and

$$dk_t = I_{\{V_t \in \partial K_0\}}\,dk_t. \tag{4.8.26}$$

Proposition 4.8.10 *The Skorokhod problem has a solution.*

Proof. Let $\tilde{W}$ be the Bellman function of our optimal control problem but with the utility function $u_\gamma = u^\gamma/\gamma$. Let us introduce the polygons

$$K_0^n := K_0 \cap \big\{n^{-2/\gamma} \le x^1 + x^2 \le n^{2/\gamma}\big\}$$

and ice-cream-shaped closed regions $\tilde{K}_0^n$ having smooth boundaries and such that $K_0^n \subseteq \tilde{K}_0^n \subseteq K_0^{n+1}$. We define on the boundary $\partial \tilde{K}_0^n$ a smooth non-tangent reflection vector field coinciding on the lateral parts of the boundary with $g(x)$. According to Theorem 5.6.3 the Skorokhod problem in each region $\tilde{K}_0^n$ admits a unique solution (V^n, k^n). Let

$$\tau^n := \inf\{t : \, |V_t^n|_1 = n^{-2/\gamma}\}, \qquad \tau := \lim \tau^n,$$
$$\rho^n := \inf\{t : \, |V_t^n|_1 = n^{2/\gamma}\}, \qquad \rho := \lim \rho^n.$$

The uniqueness of solutions allows us to assert the existence of a pair of processes (V, k) defined on the time interval $[0, \tau \wedge \rho[$ and such that (V, k) coincides with (V^n, k^n) on $[0, \tau^n \wedge \rho^n]$. From the homotheticity property it follows that the Bellman function admits the upper bound of the form

$$W(x) \le \bar{\kappa}|x|_1^\gamma, \quad x \in K,$$

and the lower bound

$$W(x) \ge \kappa|x|_1^\gamma, \quad x \in K_0.$$

Omitting in the dynamic programming inequality (4.5.4) the integral term and using afterwards the above lower bound, we obtain that

$$W(x) \ge Ee^{-\beta(\tau_n \wedge \rho_n \wedge t)} W(V_{\tau_n \wedge \rho_n \wedge t}) \ge \kappa e^{-t} E|V_{\tau_n \wedge \rho_n \wedge t}|_1^\gamma, \quad x \in K_0.$$

Since

$$E|V_{\tau_n \wedge \rho_n \wedge t}|_1^\gamma \ge EI_{\rho_n < \tau_n \wedge t}|V_{\tau_n \wedge \rho_n \wedge t}|_1^\gamma \ge n^2 P(\rho_n < \tau_n \wedge t),$$

this implies that $\sum_n P(\rho_n < \tau_n \wedge t) < \infty$. By virtue of the Borel–Cantelli lemma, $\rho_n \ge \tau_n \wedge t$ for sufficiently large n on the set of full measure. Thus, $\rho \ge \tau \wedge t$, and, because t is arbitrary, $\rho \ge \tau$.

So, we know that the processes V and k are defined on the stochastic interval $[0, \tau[$ and $\lim_n V_{\tau_n} = 0$.

One can show that $\lim_{t \uparrow \tau} V_{\tau_n} = 0$ (a.s.). In other words, the process V is absorbed at the origin. □

4.8.5 Optimal Strategy

Now we formulate the Davis–Norman theorem on the structure of the optimal solution.

Theorem 4.8.11 *Suppose that the initial endowment $x \in K_0$. Then the process V participating in the solution of the Skorokhod problem (4.8.24)–(4.8.26) defines the dynamics of the optimal portfolio, and the optimal strategy is given by the formulae*

$$B_t = \int_0^t g(V_s)\, dk_s, \tag{4.8.27}$$

$$c_t = \big(W_\xi(V_t)\big)^{1/(\gamma-1)}. \tag{4.8.28}$$

Proof. We follow the same line of arguments as in the proof of the Merton theorem. Applying the Itô formula and taking into account that $g_i W'(x) = 0$ for $x \in \partial K_0$, we obtain that

$$e^{-\beta t} W(V_t) + J_t^{\pi} = W(x) + \sigma \int_0^t V_t^2 W_\eta(V_t)\, dw_t. \tag{4.8.29}$$

The integrals with respect to dt and dk disappeared. To obtain the result, it remains to check that the stochastic integral above is a martingale (hence, its expectation is zero) and verify that, for a certain sequence of real numbers $t_n \uparrow \infty$,

$$\lim_{n\to\infty} e^{-\beta t_n} EW(V_{t_n}) = 0. \tag{4.8.30}$$

Due to the homotheticity property of the derivative of the Bellman function following from Lemma 4.7.5, we have the inequality $|W'(y)| \le \kappa |y|^{\gamma-1}$, where κ is the bound for the derivative of W on the intersection of the set K_0 with the line $\xi + \eta = 1$. Thus,

$$\big|\eta W_\eta(y)\big| \le \kappa |y|^\gamma \le \kappa\big(1 + |y|\big), \quad y \in K_0,$$

and the absolute value of the integrand is dominated by a linear function of the phase variable. Using the exponential bound of Proposition 4.2.1, we infer that the stochastic integral in (4.8.29) is a martingale.

To accomplish the proof, we need the inequality

$$W_\xi(y) \ge \kappa |y|^{\gamma-1}, \quad y \in K_0,$$

also implied by the homotheticity. Here the constant $\kappa > 0$ is the minimum of the partial derivative W_ξ on the intersection of the set K_0 with the line $\xi + \eta = 1$. It is strictly positive: the derivative of the Bellman function W in the direction g_1 is positive, the derivative in the radial direction is strictly positive, and the vector e_1 lies between these two directions.

Using these observations, we have the following chain of inequalities with a varying constant:

$$E\int_0^\infty e^{-\beta t} W(V_t)\, dt \le \kappa E \int_0^\infty e^{-\beta t} |V_t|^\gamma\, dt \le \kappa E \int_0^\infty e^{-\beta t} u(c_t)\, dt \le W(x).$$

Since W is finite, this obviously implies the existence of a sequence $t_n \uparrow \infty$ for which (4.8.30) holds. $\square$

Remark. The case where $x \in K_i$ is easily reduced to the one treated in the theorem. It is sufficient to modify the process B by adding the initial jump

$$\Delta B_0 = \inf\{s \ge 0 :\ x - s g_i \in K_0\}, \quad x \in K_i,\ i = 1, 2. \tag{4.8.31}$$

The function W on the set K_i is constant along the direction g_i for $i = 1, 2$. Thus, such a modification gives a strategy resulting in the value $W(x + \Delta_0)$

coinciding with $W(x)$. Notice that in $\operatorname{int} K_1$ (resp., in $\operatorname{int} K_2$) the changes of the initial endowment means the buying of stock (resp. the selling of stock), while in $\operatorname{int} K_0$ there are no transactions. This explains the abbreviations BS, SS, and NT used in the literature for the corresponding regions.

Using the structure of optimal control, we improve a bit Proposition 4.3.5, which gives us an upper bound for the Bellman function in the case where the solution of the classical Merton problem is finite. It happens that, "usually," in the model with transaction costs the bound is strict, and this fact plays an important role to locate more precisely the boundaries of the no-transaction cone K_0 (see the next subsection). The precise statement is as follows.

Proposition 4.8.12 *Suppose that $\kappa_M > 0$. Let $p = (p_1, p_2) \in K^*$ and $p_1 = 1$. If $(1-\gamma)\sigma^2 \neq \mu$, then*

$$W(x) < \mathbf{m}u(px) = \frac{1}{\gamma}\kappa_M^{\gamma-1}(px)^\gamma \quad \forall x \in \operatorname{int} K. \tag{4.8.32}$$

If $(1-\gamma)\sigma^2 = \mu$, then $e_2 \in K_0$.

Proof. Proposition 4.3.5 says that the function $\varphi(x) = \mathbf{m}u(px)$ is a supersolution of the HJB equation, and, therefore,

$$\mathcal{L}_0\varphi + u^*(\varphi_\xi) = -\frac{\gamma}{2(1-\gamma)\sigma^2}\left[\frac{(1-\gamma)\sigma^2 p_2\eta}{\xi + p_2\eta} - \mu\right]^2 \phi(x) \le 0.$$

The equality holds if and only if

$$\mu\xi - \left[(1-\gamma)\sigma^2 - \mu\right]p_2\eta = 0. \tag{4.8.33}$$

Let us plug-in the optimal process $V = (V^1, V^2)$ (corresponding to the strategy given by (4.8.27) and (4.8.28), eventually with an initial transfer) into the function φ. Applying the Itô formula, we get

$$\begin{aligned}
e^{-\beta t}\varphi(V_t) &= \varphi(V_0) + \int_{]0,t]} e^{-\beta s}\varphi_\eta(V_s)\sigma V_s^2\, dw_s + \int_{]0,t]} e^{-\beta s}\varphi'(V_s)g(V_s)\, dk_s \\
&\quad + \int_{]0,t]} e^{-\beta s}\left[\mathcal{L}_0\varphi(V_s) - c_s\varphi_\xi(V_s) + u(c_s)\right] ds - \int_{]0,t]} e^{-\beta s}u(c_s)\, ds \\
&\le \varphi(x) + \int_{]0,t]} e^{-\beta s}\varphi_\eta(V_s)\sigma V_s^2\, dw_s \\
&\quad + \int_{]0,t]} e^{-\beta s}\left[\mathcal{L}_0\varphi(V_s) + u^*\big(\varphi_\xi(V_s)\big)\right] ds - \int_{]0,t]} e^{-\beta s}u(c_s)\, ds.
\end{aligned}$$

The above bound holds because $\varphi(V_0) \le \varphi(x)$ due to losses which occur at the initial transfer, $g_i\varphi'(x) \le 0$ for $x \in \partial K_0$, and

$$-c_s\varphi_\xi(V_s) + u(c_s) \le u^*\big(\varphi_\xi(V_s)\big).$$

By the same arguments as in the previous proof, we infer that the expectation of the stochastic integral is zero and $Ee^{-\beta t}\varphi(V_t) \to 0$ as $t \to \infty$. It follows that

$$W(x) \le \varphi(x) + E \int_{]0,t]} e^{-\beta s}\big[\mathcal{L}_0\varphi(V_s) + u^*\big(\varphi_\xi(V_s)\big)\big]\,ds \le \varphi(x).$$

In the case $(1-\gamma)\sigma^2 \ne \mu$, the second inequality is always strict: otherwise the integrant is a negligible process, i.e., according to (4.8.33), we would have

$$\mu V^1 - \big[(1-\gamma)\sigma^2 - \mu\big]p_2 V^2 = 0.$$

This identity is impossible because the left-side is a semimartingale with non-trivial diffusion component.

If $(1-\gamma)\sigma^2 = \mu$, we have, necessarily, that $V_t^1 = 0$ for all $t > 0$. Thus, the process V after the initial jump evolves along the axis of ordinates, and therefore $e_2 \in K_0$. □

4.8.6 Precisions on the No-Transaction Region

We established already that the no-transaction region $K_0 = \text{cone}\{\tilde{g}_1, \tilde{g}_2\}$ has a nonempty interior and lies strictly above the axis of abscises. Now we give some bounds on the position of the generator $\tilde{g}_2$.

The following simple lemma ensures us that, whatever is $\lambda > 0$, the interval $[z_1, z_2]$ (depending on λ) lies inside the fixed interval, namely, $[0, 2\theta+1]$, $\theta = \beta/(1-\gamma)$. We shall need this fact for the asymptotic analysis as $\lambda \to 0$.

Lemma 4.8.13 *We have*

$$0 < z_2 \le 1 + \frac{2\theta}{1+\lambda+2\theta\lambda}. \tag{4.8.34}$$

Proof. The first inequality holds because $z_2 > z_1 > 0$. According to Proposition 4.8.2, we have the inclusion $K_2 \subseteq \text{cone}\{g_2, Ng_2 + e_2\}$, where the constant $N = 2\theta(1+\lambda)^{-1}$. The ray generated by $Ng_2 + e_2$ intersects the line $\{(\xi,\eta) :\ \xi + \eta = 1\}$ at the point $(1-z, z)$, where

$$z = 1 + \frac{N}{1+N\lambda} = 1 + \frac{2\theta}{1+\lambda+2\theta\lambda}.$$

Since the point $(1-z_2, z_2)$ is the intersection of the boundary ray separating K_2 and K_0 with the aforementioned line, we have from obvious geometric considerations that $z_2 \le z$, which is exactly the second inequality. □

With minor efforts, we can get a more precise information about the positions of z_2 and z_1.

Recall that, according to Corollary 4.8.3, on the cones K_i, $i = 1, 2$, the Bellman function W has the form $W(x) = a_i(p_i x)^\gamma$, where

$$a_1 \geq \kappa_*^{\gamma-1}, \qquad a_2 \geq \kappa_*^{\gamma-1} \frac{1}{(1+\lambda)^\gamma},$$

with $\kappa_* = \beta/(1-\gamma)$. Now we are able to say a bit more: the inequality for a_2 is strict! Indeed, suppose that we have the equality for a_2. In virtue of (4.7.7), for such a value of a_2, the function $W(x)$ in the upper half-plane dominates the function $a_2(p_2 x)^\gamma$. But we know that the latter dominates $W(x)$ on the whole cone K and coincides with $W(x)$ exactly on the cone K_2. Combining these facts, we get that $K_2 \supseteq K \cap \{x = (\xi, \eta) : \eta \geq 0\}$, which is impossible.

Now we give sharper bounds for z_i.

Proposition 4.8.14 (a) *We always have the inequality*

$$z_2 < \frac{\mu(1+\lambda)}{\frac{1}{2}(1-\gamma)\sigma^2 + \mu\lambda}. \tag{4.8.35}$$

(b) *If $\kappa_M > 0$ and $\theta \neq 1$ (i.e., $(1-\gamma)\sigma^2 \neq \mu$), then*

$$z_2 > \frac{\mu(1+\lambda)}{(1-\gamma)\sigma^2 + \mu\lambda}. \tag{4.8.36}$$

(c) *If $\kappa_M > 0$ and $\theta = 1$, then $z_2 = 1$.*
(d) *If $\kappa_M > 0$ and $(1-\gamma)\sigma^2 > \lambda\mu$, then*

$$z_1 < \frac{\mu\lambda}{(1-\gamma)\sigma^2(1+\lambda) - \mu\lambda}. \tag{4.8.37}$$

Proof. (a) Let us consider the function $\psi(x) = a_2 u(p_2 x)$, which is the solution of the HJB equation in K_2. Since $a_2 > \kappa_*^{\gamma-1}(1+\lambda)^{-\gamma}$, we have, for every $x \in \operatorname{int} K$, the strict inequality $u^*(\psi_\xi(x)) - \beta\psi(x) < 0$ and, therefore, the bound

$$\mathcal{L}_0\psi(x) + u^*\big(\psi_\xi(x)\big) < a_2(p_2 x)^{\gamma-2}\eta\left[\frac{1}{2}\sigma^2(\gamma-1)\eta + \mu(p_2 x)\right].$$

The expression in the square bracket is less or equal to zero when the point $x = (1-z, z) \in K$ and z dominates the right-hand side of (4.8.35). This observation makes the bound (4.8.35) obvious.

(b) Let us examine the right-hand side of the identity

$$\begin{aligned}
&\mathcal{L}_0\psi(x) + u^*\big(\psi_\xi(x)\big) \\
&\quad = -\frac{\gamma}{2(1-\gamma)\sigma^2}\left[\frac{(1-\gamma)\sigma^2\eta}{(1+\lambda)\xi + \eta} - \mu\right]^2 \psi(x) \\
&\qquad + \left(\frac{1}{2}\frac{\gamma}{1-\gamma}\frac{\mu^2}{\sigma^2} - \beta + (1-\gamma)a_2^{1/(\gamma-1)}(1+\lambda)^{\frac{\gamma}{\gamma-1}}\right)\psi(x).
\end{aligned}$$

In virtue of Proposition 4.8.12, under the assumed hypotheses, the coefficient $(\dots)$ is strictly positive. Thus, the function given by the expression $[\dots]$ cannot vanish at points of the set $\operatorname{int} K_2$ where the function ψ is the solution of the HJB equation. It has the positive sign on this set (its values tend to $+\infty$ as $x = (\xi, \eta)$ approaches a point of the outer boundary of K_2 other than zero). Moreover, the continuity considerations imply that $[\dots]$ is strictly positive also on the inner boundary (of course, except the origin), in particular, at the point $(1 - z_2, z_2)$. This last property is equivalent to inequality (4.8.36).

(c) In this case the Proposition 4.8.5 says that the coefficient $(\dots) = 0$ and $z_2 \le 1$. On the other hand, according to Proposition 4.8.12, $z_2 \le 1$.

(d) If $(1-\gamma)\sigma^2 = \mu$, inequality (4.8.37) is reduced to $z_1 < 1$. However, we already know that $z_1 < z_2$ always and $z_2 = 1$ in the considered case as we just proved. Suppose that $(1-\gamma)\sigma^2 \ne \mu$. For $\varphi(x) = a_1 u(p_1(x))$, we have the identity

$$\begin{aligned}\mathcal{L}_0\psi(x) + u^*\big(\psi_\xi(x)\big) = &-\frac{\gamma}{2(1-\gamma)\sigma^2}\left[\frac{(1-\gamma)\sigma^2(1+\lambda)\eta}{\xi + (1+\lambda)\eta} - \mu\right]^2 \varphi(x) \\ &+ \left(\frac{1}{2}\frac{\gamma}{1-\gamma}\frac{\mu^2}{\sigma^2} - \beta + (1-\gamma)a_1^{\frac{1}{\gamma-1}}\right)\varphi(x).\end{aligned}$$

Proposition 4.8.12 provides us the information that the second term is strictly positive on $\operatorname{int} K_1$, and we derive the required inequality (4.8.37) by the same arguments as in (a). $\square$

4.9 Liquidity Premium

4.9.1 Non-Robustness with Respect to Transaction Costs

According to Theorem 4.1.1, in the Merton two-asset model of frictionless financial market, the optimal expected utility of the unit wealth invested in a portfolio is given by the formula

$$\mathbf{m} = W_M(1) = \kappa_M^{\gamma-1}/\gamma,$$

where

$$\kappa_M := \frac{1}{1-\gamma}\left(\beta - \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{\mu^2}{\sigma^2}\right); \tag{4.9.1}$$

it is assumed that the model parameters are such that $\kappa_M > 0$. The optimal strategy prescribes to rebalance continuously the portfolio to keep the proportion of stock to the total value at the constant level $\theta = (1-\gamma)^{-1}\mu/\sigma^2$, called in the literature the *Merton proportion*.

It is natural to expect that in the market with friction, even following an appropriate optimal strategy, the investor cannot achieve the above performance. It would be interesting to know to which extent the presence of transaction costs deteriorate the portfolio performance. Unfortunately, the solution of the Davis–Norman problem does not admit such an explicit expression, and the comparison between two results seems to be complicated. It can be done asymptotically for small transaction costs. Assume that the initial endowment of the investor is in θ units of stock and $1-\theta$ units on money, where θ is the Merton proportion. The next theorem due to Shreve asserts that the discrepancy between the two optimal values, in general, is of (exact) order $\lambda^{2/3}$ as the transaction cost coefficient λ tends to zero. Thus, the model is not robust in the sense that the discrepancy increases infinitely fast when the transaction costs appear. The only exception is the case $\theta = 1$, where the portfolio has zero position in money, and the stock is sold only to consume. This case will be considered separately at the end of the section.

In our presentation, in order to have simpler formulae, we assume that both operations, buying stock and selling stock, are charged equally.

Theorem 4.9.1 *Suppose that $\theta \neq 1$. Then there are constants κ_1, $\kappa_2 > 0$, independent of λ, such that*

$$\mathbf{m} - \kappa_2 \lambda^{2/3} \le W(1-\theta, \theta) \le \mathbf{m} - \kappa_1 \lambda^{2/3} \tag{4.9.2}$$

for all sufficiently small $\lambda > 0$.

Proof. Recall that the function $\psi(z) := W(1-z, z)$ is concave, continuously differentiable on the interval $]-1/\lambda, 1+1/\lambda[$, and its second derivative may have (jump) discontinuities only at two (distinct) points z_1, z_2. It satisfies, in the classical sense, everywhere except at these two points, the HJB equation

$$\max\{\ell_1\psi(z), \ell_0\psi(z), \ell_2\psi(z)\} = 0 \tag{4.9.3}$$

involving the first-order operators

$$\ell_1\psi(z) = -\lambda\gamma\psi(z) + (1+\lambda z)\psi'(z), \qquad \ell_2\psi(z) = -\lambda\gamma\psi - (1+\lambda-\lambda z)\psi'(z),$$

and the second-order operator

$$\ell_0\psi(z) = f_2(z)\psi''(z) + f_1(z)\psi'(z) + f_0(z)\psi(z) + \frac{1-\gamma}{\gamma}\left[\gamma\psi(z) - z\psi'(z)\right]^{\frac{\gamma}{\gamma-1}}$$

with the coefficients

$$\begin{aligned} f_2(z) &= \frac{1}{2}\sigma^2 z^2 (1-z)^2, \\ f_1(z) &= -\sigma^2(1-\gamma)z(1-z)(z-\theta), \\ f_0(z) &= -\frac{1}{2}\sigma^2\gamma(1-\gamma)(z-\theta)^2 - (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}. \end{aligned}$$

The structure of ψ is as follows. Outside of the interval $]z_1, z_2[$, the function $\psi^{1/\gamma}$ coincides with two linear functions: on $[-1/\lambda, z_1]$ with a function proportional to $(1+\lambda z)$, thus, vanishing at $-1/\lambda$; on $[z_2, 1+1/\lambda]$ with a function proportional to $(1+\lambda-\lambda z)$, thus, vanishing at $1+1/\lambda$. On the interval $]z_1,z_2[$ the function ψ is the classical solution of the second-order differential equation $\ell_0\psi(z)=0$.

For a fixed λ, the graph of ψ is an arc-shaped curve, flattening and with the increasing base as λ tends to zero. We may expect that the "interpolation" interval degenerates into the single point θ and that the maximal value of ψ converges to $\mathbf{m}$.

Having in mind this behavior, we approximate ψ from above and below by C^1-functions $\tilde\psi$ having a similar shape. We interpolate their "linear in the power γ" parts by parabolas and choose, for these functions, the interpolation intervals $]\tilde z_1,\tilde z_2[$ containing θ and being of the length of order $\lambda^{1/3}$.

Namely, for $r>0$, we put

$$Q(z):=\mathbf{m}-r\lambda^{2/3}-(z-\theta)^2\lambda^{2/3},$$

$\tilde z_1:=\theta-\delta_1\lambda^{1/3}$, and $\tilde z_2:=\theta+\delta_2\lambda^{1/3}$, where the bounded strictly positive coefficients $\delta_i=\delta_i(\lambda;r)$ will be chosen to guarantee the continuity of the first derivatives of the function $\tilde\psi(z)=\tilde\psi(z,\lambda;r)$. The latter is defined (for sufficiently small λ) by the formula

$$\begin{aligned}\tilde\psi(z):=&\frac{Q(\tilde z_1)}{(1+\lambda\tilde z_1)^\gamma}(1+\lambda z)^\gamma I_{[-1/\lambda,\tilde z_1[}(z)+Q(z)I_{[\tilde z_1,\tilde z_2]}(z)\\&+\frac{Q(\tilde z_2)}{(1+\lambda-\lambda\tilde z_2)^\gamma}(1+\lambda-\lambda z)^\gamma I_{]\tilde z_2,1+1/\lambda]}(z).\end{aligned}\qquad(4.9.4)$$

Its first derivative:

$$\tilde\psi'(z)=\begin{cases}\frac{\gamma\lambda}{1+\lambda z}\tilde\psi(z), & z\in]-1/\lambda,\tilde z_1[,\\ -2\lambda^{2/3}(z-\theta), & z\in]\tilde z_1,\tilde z_2[,\\ -\frac{\gamma\lambda}{1+\lambda-\lambda z}\tilde\psi(z), & z\in]\tilde z_2,1+1/\lambda[.\end{cases}$$

Its second derivative:

$$\tilde\psi''(z)=\begin{cases}-\frac{\gamma(1-\gamma)\lambda^2}{(1+\lambda z)^2}\tilde\psi(z), & z\in]-1/\lambda,\tilde z_1[,\\ -2\lambda^{2/3}, & z\in]\tilde z_1,\tilde z_2[,\\ -\frac{\gamma(1-\gamma)\lambda^2}{(1+\lambda-\lambda z)^2}\tilde\psi(z), & z\in]\tilde z_2,1+1/\lambda[.\end{cases}$$

The second derivative is constant over the central interval. It is continuous on the external intervals, and its limits at the points $\tilde z_i$ coincide with the corresponding one-sided derivatives.

The result will be established if we find, independently on λ, two values of the parameter $r>0$ such that the corresponding $\delta_i(\lambda)$ belongs to a bounded interval, $\tilde\psi(z,\lambda;r)$ is a supersolution of the HJB equation for the smaller value (say, r_1) and a subsolution for the larger value (say, r_2) whatever are $\lambda \le \lambda_0$; the threshold λ_0 may depend on r_i. Indeed, in virtue of the comparison lemma given after the proof of the theorem, the function ψ will lie between these two functions, and the corresponding inequalities for the values calculated at the point θ yield (4.9.2). Note that the supersolution and the subsolution are understood in an "almost" classical sense: the corresponding inequalities for $\max_i \ell\tilde{\psi}(z)$ should hold everywhere except, maybe, at the points $\tilde{z}_i$.

We have

$$\begin{aligned}
\tilde{\psi}(\tilde{z}_i) &= \mathbf{m} - r\lambda^{2/3} - \delta_i^2\lambda^{4/3}, \quad i = 1, 2,\\
D^+\tilde{\psi}(z_1) &= 2\delta_1\lambda, \qquad D^-\tilde{\psi}(z_2) = -2\delta_2,\\
D^-\tilde{\psi}(z_1) &= \frac{\gamma\lambda}{1+\lambda\tilde{z}_1}Q(\tilde{z}_1), \qquad D^+\tilde{\psi}(z_2) = -\frac{\gamma\lambda}{1+\lambda-\lambda\tilde{z}_2}Q(\tilde{z}_2).
\end{aligned}$$

The requirement that $\tilde{\psi}$ is continuously differentiable means that the right and left derivatives of $\tilde{\psi}$ at each point $\tilde{z}_i$ coincide. It is met when δ_i are the (positive) roots of the corresponding quadratic equations

$$A_i(\lambda)\delta^2 - B_i(\lambda)\delta + C_i(\lambda) = 0.$$

Their coefficients are as follows:

$$\begin{aligned}
A_1(\lambda) &= (2-\gamma)\lambda^{4/3}, \qquad B_1(\lambda) = 2(1+\lambda\theta), \qquad C_1(\lambda) = \gamma\big[\mathbf{m} - r\lambda^{2/3}\big];\\
A_2(\lambda) &= (2+\gamma)\lambda^{4/3}, \qquad B_2(\lambda) = 2(1+\lambda-\lambda\theta), \qquad C_2(\lambda) = C_1(\lambda).
\end{aligned}$$

Note that, for λ close to zero, the left-hand sides of these equations are positive for $\delta = B_i(\lambda)/C_i(\lambda)$ and negative for $\delta = 2B_i(\lambda)/C_i(\lambda)$. Thus, we can find roots $\delta_i(\lambda)$ in this interval. Asymptotically,

$$\delta_i(\lambda) \in \big[2/(\gamma\mathbf{m}), 4/(\gamma\mathbf{m}) + o(1)\big], \tag{4.9.5}$$

and, therefore, $\delta_i(\lambda) \in [2/(\gamma\mathbf{m}), 5/(\gamma\mathbf{m})]$ for λ less than a certain λ_0 depending on r.

Let us check that, for "r small" (respectively, "r large"), the function $\tilde{\psi}$ is a supersolution (respectively, subsolution) of the HJB equation (4.9.3).

In principle, we have to verify 18 inequalities. Luckily, most of them are trivial or easy. We organize our analysis in three parts.

1. The interval $]-1/\lambda, \tilde{z}_1[$.

We have here $\ell_1\tilde{\psi} = 0$, and, hence, the subsolution inequality ≥ 0 is obvious. Moreover, since $\tilde{\psi}$ and $\tilde{\psi}'$ are both positive, the inequality $\ell_2\tilde{\psi} \le 0$ always holds. To check the supersolution inequality ≤ 0, it remains to verify, on the considered interval, that $\ell_0\tilde{\psi} \le 0$.

Note that the function $\tilde{\psi}(z)$ on the considered interval is proportional to $(1+\lambda z)^{\gamma}$, and, therefore,

$$\gamma\tilde{\psi}(z) - z\tilde{\psi}'(z) = \gamma\frac{1}{1+\lambda z}\tilde{\psi}(z) = \text{const } \times (1+\lambda z)^{\gamma-1}.$$

Since $u^*(p) = (1-\gamma)/\gamma p^{\gamma/(\gamma-1)}$, the nonlinear term is proportional to $\tilde{\psi}$, i.e.,

$$u^*\big(\gamma\tilde{\psi}(z) - z\tilde{\psi}'(z)\big) = \kappa_\lambda\tilde{\psi}(z).$$

Since $\tilde{\psi}(\tilde{z}_1) = Q(\tilde{z}_1)$, we can determine the constant κ_λ. Namely,

$$\kappa_\lambda = (1+\lambda\tilde{z}_1)^{\frac{\gamma}{1-\gamma}}(1-\gamma)\gamma^{\frac{1}{\gamma-1}}\big[Q(\tilde{z}_1)\big]^{\frac{1}{\gamma-1}}.$$

Our definitions imply that $\tilde{z}_1 = \theta + o(1)$, $Q(\tilde{z}_1) = \mathbf{m} - r\lambda^{2/3} + o(\lambda^{2/3})$. The derivative $Q'(1) = 1$, and, hence, κ_λ has the following asymptotic expansion:

$$\kappa_\lambda = (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\left(1 + \frac{1}{1-\gamma}\frac{r}{\mathbf{m}}\lambda^{2/3}\right) + o\big(\lambda^{2/3}\big).$$

Inspecting the function $\ell_0\tilde{\psi}$, we see that the term $f_2(z)\tilde{\psi}''(z)$ is always negative. If $z \le 0$, the coefficient f_1 is negative, and so is $f_1(z)\tilde{\psi}'(z)$. Omitting these two terms and taking into account that, for $z \in]-1/\lambda, \tilde{z}_1[$,

$$\begin{aligned} f_0(z) &= -\frac{1}{2}\sigma^2\gamma(1-\gamma)(z-\theta)^2 - (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}} \\ &\le -\frac{1}{2}\sigma^2\gamma(1-\gamma)\delta_1^2\lambda^{2/3} - (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}, \end{aligned}$$

we arrive, on the interval $]-1/\lambda, 0]$, at the bound

$$\frac{\ell_0\tilde{\psi}(z)}{\tilde{\psi}(z)} \le f_0(z) + \kappa_\lambda \le \frac{1}{2}\sigma^2\gamma(1-\gamma)\left(\kappa r - \frac{1}{(\gamma\mathbf{m})^2}\right)\lambda^{2/3} + o\big(\lambda^{2/3}\big),$$

which holds for all $\lambda \le \lambda_0$ (the threshold is chosen to insure that there exists a positive root $\delta_1 \ge 1/(\gamma\mathbf{m})$). Here the constant $\kappa > 0$, and, hence, for "small" r, the coefficient of the main term is strictly negative.

On the interval $]0, \tilde{z}_1[$ we cannot omit the term with the first derivative, but this does not affect the resulting asymptotic expansion: on this interval, $f_1(z)$ is bounded, and

$$f_1(z)\frac{\tilde{\psi}'(z)}{\tilde{\psi}(z)} = f_1(z)\frac{\gamma\lambda}{1+\lambda z} = o(\lambda)$$

uniformly in z.

2. The interval $]\tilde{z}_2, 1+1/\lambda[$.

The reasoning is completely analogous to the one given above.

3. The interval $]\tilde{z}_1, \tilde{z}_2[$.

Here the function

$$\ell_1\tilde{\psi}(z) = \ell_1 Q(z) = -\lambda\gamma Q(z) + (1+\lambda z)Q'(z)$$

is quadratic in z, and its coefficients at z^2 and z are, respectively, $\lambda^{5/3}(\gamma - 2)$ and $2\lambda^{5/3}(-\gamma\theta + \theta - 1/\lambda)$. The point

$$z_{\max}(\lambda) := \frac{\theta(1-\gamma) - 1/\lambda}{2-\gamma},$$

where $\ell_1 Q(z)$ attains its maximum, tends to $-\infty$ as $\lambda \to 0$ and, hence, lies leftwards to the interval $[\tilde{z}_1, z_2]$ for sufficiently small λ. But this means that, on this interval, $\ell_1\tilde{\psi}$ decreases, i.e., $\ell_1\tilde{\psi}(z) \le \ell_1\tilde{\psi}(\tilde{z}_1) = 0$.

Exactly by the same arguments we check that $\ell_2\tilde{\psi}(z) \le 0$.

Summarizing: in the central interval, the super- or subsolution property for $\tilde{\psi}$ is equivalent, respectively, to the inequality $\ell_0\tilde{\psi} \le 0$ or $\ell_0\tilde{\psi} \ge 0$.

On the considered interval (degenerating in the limit to the point θ) we have that $Q(z) = \mathbf{m} - r\lambda^{2/3} + o(\lambda^{2/3})$, $Q'(z) = o(\lambda^{2/3})$ uniformly in z and $Q'' = -2\lambda^{2/3}$. With this, we get readily that

$$\ell_0\tilde{\psi}(z) = -\sigma^2\theta(1-\theta)\lambda^{2/3} + r\frac{1}{2}\sigma^2\gamma(1-\gamma)(z-\theta)^2 + r\kappa\lambda^{2/3} + o\big(\lambda^{2/3}\big)$$

with some constant $\kappa > 0$. Since $(z-\theta)^2 \le 5(\gamma\mathbf{m})\lambda^{2/3}$, this representation makes clear that in the case $\theta \ne 1$ we can choose a small value r such that $\ell_0\tilde{\psi} \le 0$ for sufficiently small λ. The opposite inequality for large values of r holds obviously for any θ. □

Remark. The condition $\theta \ne 1$ is needed only at the very end of the proof. Moreover, the subsolution property of the approximation is not violated even in the exceptional case, and, therefore, due to the comparison lemma below, the left inequality in (4.9.2) remains valid. In contrast to this, for $\theta = 1$, the right inequality fails because, as we show later, in this case the difference $\psi(\theta) - \mathbf{m}$ converges to zero with rate λ.

Now we are back to the comparison lemma. It is extremely simple for the C^1-functions which are twice continuously differentiable everywhere except a finite number of points where the limits of the second derivatives exist and coincide with the one-sided second derivatives. A subtlety is that one of the functions to be compared is ψ, which is simultaneously a super- and subsolution, but for which we cannot guarantee the mentioned behavior of the second derivatives at the point $z = 1$ (in the case $z_2 = 1$). However, $(z_n - 1)^2\psi''(z_n) \to 0$ for certain sequences $z_n \uparrow 0$ and $z_n \downarrow 0$. But this implies that the super- and subsolution inequalities hold also at the point $z = 1$ with the degenerate operator ℓ_0 defined by

$$\ell_0\psi(1) = f_0(1)\psi(1) + u^*\big(\gamma\psi(1) - \lambda\psi'(1)\big).$$

In the formulation below we suppose that ψ_1 and ψ_2 possess this property of the second derivative at $z = 1$.

Lemma 4.9.2 *Let ψ_1 be a supersolution, and ψ_2 be a subsolution with the same boundary condition. Then $\psi_1 \geq \psi_2$.*

Proof. Let us consider a point z_0 where the difference $\psi_2 - \psi_1$ attains its maximum. If the claim fails, then $\psi_2(z_0) > \psi_1(z_0)$. Since $\psi_2'(z_0) = \psi_1'(z_0)$ and ψ_1 is a supersolution, $l_i\psi_2(z_0) < l_i\psi_1(z_0) \leq 0$ for $i = 1, 2$. Suppose first that at z_0 the second derivatives are continuous. Then $\psi_2''(z_0) \leq \psi_1''(z_0)$. Taking into account the signs of the coefficients $f_2(z_0)$ and $f_0(z_0)$ and using the fact that u^* is decreasing, we infer that $l_0\psi_2(z_0) < l_0\psi_1(z_0) \leq 0$. Thus, all $l_i\psi_2(z_0)$ are strictly less than zero, in contradiction with the subsolution property. The general case $z_0 \neq 1$ is not much different: the one-sided second derivatives of $\psi_2 - \psi_1$ are negative at z_0, and we obtain, as before, that $l_0\psi_2(z_0+) < 0$. This means that the subsolution property of ψ_2 is violated in a neighborhood of z_0. At last, if $z_0 = 1$, we arrive at a violation of the subsolution property at this point due to the remark preceding the lemma. □

4.9.2 First-Order Asymptotic Expansion

A more elaborated analysis based on the same kind of approximation, but with the interpolating polynomials of the fourth order of the form

$$\begin{aligned} Q(z) := \mathbf{m} - r_2\lambda^{2/3} - r_3\lambda - r_4\lambda^{4/3} \\ - \rho_1(z-\theta)\lambda - \rho_2(z-\theta)^2\lambda^{2/3} + \rho_3(z-\theta)^2\lambda^{1/3} - \rho_4(z-\theta)^4, \end{aligned}$$

allows us to establish for the discrepancy the exact asymptotics of order $\lambda^{2/3}$ with an explicit expression for the constant.

Theorem 4.9.3 *Suppose that $\theta \neq 1$. Then*

$$W(1-\theta, \theta) - \mathbf{m} = r_2\lambda^{2/3} + o\big(\lambda^{2/3}\big), \tag{4.9.6}$$

where

$$r_2 = \left(\frac{9}{32}(1-\gamma)\theta^4(1-\theta)^4\right)^{1/3} (\gamma\mathbf{m})^{1+\frac{1}{1-\gamma}}\sigma^2. \tag{4.9.7}$$

Proof. The arguments require a bit patience, but for the reader accustomed already with the structure of coefficients and notations, it is quite a routine. The approximating functions are again given by formula (4.9.4) but with the interpolating polynomial $Q(z)$ of the fourth order with the maximum attained at θ, namely, with

$$Q(z) := \mathbf{m} - r_2\lambda^{2/3} - r_3\lambda - \rho_2(z-\theta)^2\lambda^{2/3} - \rho_4(z-\theta)^4. \tag{4.9.8}$$

Note that the first part is a series expansion in powers of $\lambda^{1/3}$ with the coefficients r_i, the second part, more involved, is an expansion in increasing powers of $z-\theta$ with the coefficients ρ_i multiplied by decreasing powers of $\lambda^{1/3}$. The coefficients r_1, r_4, and ρ_1, ρ_3 are already taken zero. We show that a good choice for other constants is the following:

$$\rho_2 := \frac{1}{\sigma^2\theta^2(1-\theta)^2}(\gamma\mathbf{m})^{\frac{1}{\gamma-1}} r_2, \tag{4.9.9}$$

$$\rho_4 := -\frac{1}{12}\frac{\gamma(1-\gamma)}{\theta^2(1-\theta)^2}\mathbf{m}. \tag{4.9.10}$$

The parameter r_3 is "free": it serves to produce sub- and supersolutions.

The extremities of the central interval will be

$$\tilde{z}_1 = \theta - \nu\lambda^{1/3} + O\big(\lambda^{2/3}\big), \qquad \tilde{z}_2 = \theta + \nu\lambda^{1/3} + O\big(\lambda^{2/3}\big), \tag{4.9.11}$$

with the positive constant ν determined by the relation

$$\frac{1}{2}\sigma^2\gamma(1-\gamma)\nu^2 = (\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\mathbf{m}^{-1} r_2. \tag{4.9.12}$$

The parameters should be chosen to ensure that $\tilde{\psi} \in C^1$.

Our analysis will go in the inverse direction to that in the proof of the previous theorem. Though we have already listed the explicit values of the constants, we shall see in a clear and successive way how they appear to eliminate terms of lower orders.

The derivatives of $\tilde{\psi}$ are given by the formulae

$$\tilde{\psi}'(z) = \begin{cases} \frac{\gamma\lambda}{1+\lambda z}\tilde{\psi}(z), & z \in]-1/\lambda, \tilde{z}_1[, \\ -2[\rho_2(z-\theta)\lambda^{2/3} + 2\rho_4(z-\theta)^3], & z \in]\tilde{z}_1, \tilde{z}_2[, \\ -\frac{\gamma\lambda}{1+\lambda-\lambda z}\tilde{\psi}(z), & z \in]\tilde{z}_2, 1+1/\lambda[, \end{cases}$$

$$\tilde{\psi}''(z) = \begin{cases} -\frac{\gamma(1-\gamma)\lambda^2}{(1+\lambda z)^2}\tilde{\psi}(z), & z \in]-1/\lambda, \tilde{z}_1[, \\ -2[\rho_2\lambda^{2/3} + 6\rho_4(z-\theta)^2], & z \in]\tilde{z}_1, \tilde{z}_2[, \\ -\frac{\gamma(1-\gamma)\lambda^2}{(1+\lambda-\lambda z)^2}\tilde{\psi}(z), & z \in]\tilde{z}_2, 1+1/\lambda[. \end{cases}$$

First, we consider the approximations on the interpolation interval $[\tilde{z}_1, \tilde{z}_2]$ and relate constants in such a way that the sign of the inequality for $\ell_0\tilde{\psi}$ will be determined by the sign of the coefficient at λ.

Using the symbol $\approx$ to denote equalities which hold up to $O(\lambda^{4/3})$ (uniformly in z) and replacing by the symbol ... the coefficients for which explicit expressions are of no importance, we represent the asymptotic expansions for the linear terms of the operator in the following transparent form:

$$f_2(z)\tilde{\psi}''(z) \approx -\sigma^2\big[\theta^2(1-\theta)^2 + \ldots(z-\theta)\big]\big[\rho_2\lambda^{2/3} + 6\rho_4(z-\theta)^2\big],$$
$$f_1(z)\tilde{\psi}'(z) \approx 0,$$
$$f_0(z)\tilde{\psi}(z) \approx -\frac{1}{2}\sigma^2\gamma(1-\gamma)(z-\theta)^2\mathbf{m} - (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\big(\mathbf{m} - r_2\lambda^{2/3} - r_3\lambda\big).$$

Since

$$\gamma\tilde{\psi}(z) - z\tilde{\psi}'(z) \approx \gamma\big(\mathbf{m} - r_2\lambda^{2/3} - r_3\lambda\big) + \ldots (z-\theta)\lambda^{2/3},$$

the nonlinear term admits the expansion

$$u^*\big(\gamma\tilde{\psi}(z) - z\tilde{\psi}'(z)\big)$$
$$\approx \frac{1-\gamma}{\gamma}(\gamma\mathbf{m})^{\frac{\gamma}{\gamma-1}} + (\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\big(\gamma r_2\lambda^{2/3} + \gamma r_3\lambda + \ldots (z-\theta)\lambda^{2/3}\big).$$

Summing up, we obtain the approximation

$$\ell_0\tilde{\psi}(z) \approx \ldots (z-\theta)\lambda^{2/3} + \ldots (z-\theta)^3 + r_3\gamma(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\lambda;$$

the coefficients at $\lambda^{2/3}$ and $(z-\theta)^2$ are zero because of our choice of ρ_2 and ρ_4, see relations (4.9.9) and (4.9.10).

On the considered interval the width of which is controlled by the parameter ν, the absolute value of the first two terms of the right-hand side is dominated by $\kappa_\nu\lambda + o(\lambda)$, where κ_ν is a constant. This observation leads to the following important conclusion: whatever is ν, one can always find r_3 sufficiently large in absolute value such that $\ell_0\tilde{\psi} \geq 0$ or $\ell_0\tilde{\psi} \leq 0$ in dependence of whether r_3 is positive or negative (of course, for $\lambda \leq \lambda_0$). Automatically, r_3 takes the control over the sign of $\max_i \ell_i\tilde{\psi}$ because the functions $\ell_1\tilde{\psi}$ and $\ell_2\tilde{\psi}$ on the central interval are negative for small λ. The latter property follows easily from the asymptotic expansions. Indeed,

$$\ell_1\tilde{\psi}(\tilde{z}_1) = -\lambda\gamma Q(\tilde{z}_1) + (1+\lambda\tilde{z}_1)Q'(\tilde{z}_1) = -\gamma\mathbf{m} + \lambda + o(\lambda),$$

while, for the derivative, we have

$$\big[\ell_1\tilde{\psi}(z)\big]' = \lambda(1-\gamma)Q'(z) + (1+\lambda\tilde{z})Q''(z) = -2\rho_2\lambda^{2/3} + o\big(\lambda^{2/3}\big).$$

Thus, the function $\ell_1\tilde{\psi}$ is negative on $[\tilde{z}_1, \tilde{z}_2]$, being negative at the left extremity and decreasing on the whole interval. The arguments for $\ell_2\tilde{\psi}$ are exactly the same.

Let us examine now the situation on the interval $]-1/\lambda, \tilde{z}_1[$. As was explained in the previous proof, the only nontrivial part is to establish the supersolution inequality $\ell_0\tilde{\psi} \leq 0$, which we expect to hold for large negative values of r_3.

We have

$$\frac{\ell_0\tilde{\psi}(z)}{\tilde{\psi}(z)} = -f_2(z)\frac{\gamma(1-\gamma)\lambda^2}{(1+\lambda z)^2} + f_1(z)\frac{\gamma\lambda}{1+\lambda z} + f_0(z) + \kappa_\lambda, \qquad (4.9.13)$$

where the coefficient

$$\kappa_\lambda = (1+\lambda\tilde{z}_1)^{\frac{\gamma}{1-\gamma}}(1-\gamma)\gamma^{\frac{1}{\gamma-1}}\big[Q(\tilde{z}_1)\big]^{\frac{1}{\gamma-1}}$$
$$\approx (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\left[1 + \frac{1}{1-\gamma}\mathbf{m}^{-1}r_2\lambda^{2/3} + \left(\frac{1}{1-\gamma}\mathbf{m}^{-1}r_3 - \frac{\gamma}{1-\gamma}\nu\right)\lambda\right].$$

Outside the interpolation interval we have the inequality

$$f_0(z) \le -\frac{1}{2}\sigma^2\gamma(1-\gamma)\big(\nu^2\lambda^{2/3} - 2\nu\eta\lambda\big) - (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}} + O\big(\lambda^{4/3}\big),$$

where $\eta = |\eta_1| \vee |\eta_2|$.

The first term in the left-hand side of (4.9.13) is negative. On the subinterval $]-1/\lambda, 0]$ the second term is also negative, and, hence,

$$\frac{\ell_0\tilde{\psi}(z)}{\tilde{\psi}(z)} \le f_0(z) + \kappa_\lambda \le \big((\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\mathbf{m}^{-1}r_3 + \kappa\big)\lambda + o(\lambda)$$

because (4.9.12) is aimed to eliminate the coefficient at $\lambda^{2/3}$ in the right-hand side. The value of the constant κ is of no importance. On the subinterval $]0, \tilde{z}_1[$, where we can affirm only that

$$\frac{\ell_0\tilde{\psi}(z)}{\tilde{\psi}(z)} \le f_1(z)\frac{\gamma\lambda}{1+\lambda z} + f_0(z) + \kappa_\lambda;$$

the structure of the resulting estimate remains the same since here the coefficient $f_1(z)$ is bounded. With this, it is clear that $\ell_0\tilde{\psi} \le 0$ for large negative values of the parameter r_3 whatever is λ smaller than some threshold value.

The situation on the other external interval is exactly the same.

Until now we did not need any specific value of r_2. It appears from the conditions of the C^1-fit at the points $\tilde{z}_i$, which are as follows:

$$\frac{\gamma\lambda}{1+\lambda\tilde{z}_1}Q(\tilde{z}_1) + 2\rho_2(\tilde{z}_1-\theta)\lambda^{2/3} + 4\rho_4(\tilde{z}_1-\theta)^3 = 0,$$
$$\frac{\gamma\lambda}{1+\lambda-\lambda\tilde{z}_2}Q(\tilde{z}_2) - 2\rho_2(\tilde{z}_2-\theta)\lambda^{2/3} - 4\rho_4(\tilde{z}_2-\theta)^3 = 0.$$

Take formally $\tilde{z}_i = \nu\lambda^{1/3} + \eta\lambda^{2/3}$ and consider the asymptotic expansions of the left-hand sides of these relations (denoted by $F_i(\tilde{z}_i)$) in powers of $\lambda^{1/3}$. Equating the coefficients at λ, we obtain the same relation for both identities:

$$\gamma\mathbf{m} - 2\rho_2\nu - 4\rho_4\nu^3 = 0.$$

In virtue of (4.9.9) and (4.9.12), the coefficient ρ_2 is proportional to ν^2, and, therefore, this is a linear equation for ν^3. Its solution is

$$\nu^3 = \frac{3}{2}\frac{\theta^2(1-\theta)^2}{1-\gamma}.$$

Expressing r_2 from (4.9.12), we arrive at formula (4.9.7) for the coefficient r_2 in the formulation of the theorem.

Examine, e.g., the case $i = 1$. Clearly, the coefficient at $\lambda^{4/3}$ is a (nondegenerate) linear function of η_1, which vanishes at some value η_1^0. Due to our choice of ν, this coefficient determines, for sufficiently small λ, the signs of $F_1(\nu\lambda^{1/3} + (\eta_1^0 + 1)\lambda^{1/3})$ and $F_1(\nu\lambda^{1/3} + (\eta_1^0 - 1)\lambda^{1/3})$. Since they are opposite, the continuity implies that there is $\eta_1(\varepsilon) \in [\eta_1^0 - 1, \eta_1^0 + 1]$ such that $F_1(\nu\lambda^{1/3} + \eta_1(\varepsilon)\lambda^{1/3}) = 0$. Thus, we established the existence of the points $\tilde{z}_i$ satisfying (4.9.11) and ensuring the smooth fit of the interpolation.

The proof is completed. □

In the above reasoning we have proved a bit more than it was claimed in the formulation of the theorem. Namely, we have shown that ψ lies between two arch-shaped functions $\tilde{\psi}_1$ and $\tilde{\psi}_2$ depending on the parameter λ and converging to each other uniformly with rate λ. With this, we can easily get the asymptotics of the extremities z_i of the no-transaction interval. The following theorem asserts that, asymptotically, the length of the latter is proportional to $\lambda^{1/3}$. The no-transaction region opens wider very quickly with the introduction of transaction costs!

Theorem 4.9.4 *Suppose that $\theta \neq 1$. Then*

$$z_1 = \theta - \nu\lambda^{1/3} + O\big(\lambda^{2/3}\big), \qquad z_2 = \theta - \nu\lambda^{1/3} + O\big(\lambda^{2/3}\big). \tag{4.9.14}$$

Proof. Recall that $z_1 > 0$ and z_2 is bounded from above, i.e., $[z_1, z_2]$ lies in a certain fixed interval $[\zeta_1, \zeta_2]$ containing θ and not depending on λ. The derivative of the concave function ψ on this fixed interval decreases from its maximal (positive) value at ζ_1 to its minimal (negative) value at ζ_2 The interval $[\tilde{z}_1, \tilde{z}_2]$ shrinks to θ. For sufficiently small λ, the functions $\tilde{\psi}_1^{1/\gamma}$ and $\tilde{\psi}_2^{1/\gamma}$ on both intervals external to $[\zeta_1, \zeta_2]$ are linear, as well as the function $\psi^{1/\gamma}$ lying between them. Taking into account that the derivatives of $\tilde{\psi}_i^{1/\gamma}$ at ζ_1 and ζ_2 are of order λ, we conclude that

$$\sup_{z\in[\zeta_1,\zeta_2]} \big|\psi'(z)\big| = O(\lambda).$$

On the interval $[\zeta_1, \zeta_2]$,

$$u^*\big(\gamma\psi(z) - z\psi'(z)\big) = \frac{1-\gamma}{\gamma}(\gamma\mathbf{m})^{\frac{\gamma}{\gamma-1}} + (\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\gamma r_2\lambda^{2/3} + O(\lambda).$$

Clearly,

$$f_0(z_i)\psi(z_i) = -\frac{1}{2}\sigma^2\gamma(1-\gamma)\mathbf{m}(z_i-\theta)^2 - (1-\gamma)(\gamma\mathbf{m})^{\frac{1}{\gamma-1}}\big(\mathbf{m} - r_2\lambda^{2/3}\big) + O(\lambda).$$

Recall that the equation $\ell_0\psi(z_i) = 0$ is always fulfilled under the convention that the term with the second derivative (may be not existing) is omitted at $z_i = 1$. In the case $z_i \neq 1$ the second derivative of ψ does exist. Since ψ is given by explicit formulae to the left of z_1 and to the right of z_2, we get easily that $\psi''(z_i) = O(\lambda^2)$. Thus, the terms with the second and first derivatives are negligible, and

$$\begin{aligned}\ell_0\psi(z_i) &= f_0(z_i)\psi(z_i) + u^*\big(\gamma\psi(z_i) - z\psi'(z_i)\big) + O(\lambda) \\ &= -\frac{1}{2}\sigma^2\gamma(1-\gamma)\mathbf{m}(z_i-\theta)^2 + (\gamma\mathbf{m})^{\frac{1}{\gamma-1}} r_2\lambda^{2/3} + O(\lambda).\end{aligned}$$

Expressing r_2 via ν from (4.9.12) and equating the left-hand side to zero, we obtain that, necessarily,

$$(z_i-\theta)^2 = \nu^2\lambda^{2/3} + O(\lambda) = \nu^2\lambda^{2/3}\big(1 + O\big(\lambda^{1/3}\big)\big)$$

and, therefore,

$$z_i = \theta \pm \nu\lambda^{1/3} + O\big(\lambda^{2/3}\big)$$

with minus for z_1 and plus for z_2. □

4.9.3 Exceptional Case: $\theta = 1$

We consider now a very particular situation where $\theta = 1$ and the arguments of the previous subsection cannot be used.

However, according to Proposition 4.8.5, for this case we have in the cone $K \cap \{x = (\xi,\eta) : \ \xi \le 0\}$ the explicit formula

$$W(x) = \mathbf{m}\frac{1}{(1+\lambda)^\gamma}(p_2x)^\gamma,$$

and, therefore,

$$\psi(\theta) = \psi(1) = W(e_2) = \mathbf{m}\frac{1}{(1+\lambda)^\gamma} = \mathbf{m} - \mathbf{m}\gamma\lambda + o(\lambda).$$

Moreover, $z_2 = 1$. For z_1, the arguments of the proof of the above theorem can be repeated with $r_2 = 0$ until the last step. Its appropriate modification shows that $z_1 = 1 + O(\lambda^{1/2})$.