Duality in constrained optimal investment and consumption problems: a synthesis

L.C.G. Rogers[1]

University of Cambridge

1 Dual problems made easy

These lectures are all about optimal investment/consumption problems, usually with some 'imperfection', such as transaction costs, or constraints on the permitted portfolios, or different interest rates for borrowing and lending, or margin requirements for borrowing, or even just incomplete markets. Some time ago, Karatzas, Lehoczky & Shreve (1987), and Cox & Huang (1989) realised that the use of duality methods provided powerful insights into the solutions of such problems, using them to prove the form of the optimal solution to significant generalisations of the original Merton (1969) problem, which Merton had proved using (very problem-specific) optimal control methods. These duality methods have become very popular in the intervening years, and now it seems that when faced with one of these problems, the steps are:

(i) try to solve the problem explicitly;

(ii) if that fails, find the dual form of the problem;

(iii) try to solve the dual problem;

(iv) if that fails, assume that investors have log utilities and try (iii) again;

(v) if that still fails, generalize the problem out of recognition and redo (ii);

(vi) write a long and technical paper, and submit to a serious journal.

As so often happens, when all the details are written up, it can be very hard to

[1]Statistical Laboratory, Wilberforce Road, Cambridge CB3 0WB, UK (phone = +44 1223 766806, fax = +44 1223 337956, e-mail = L.C.G.Rogers@statslab.cam.ac.uk). Supported partly by EPSRC grant GR/R03006. It is a pleasure to thank my hosts at the Workshop on Financial Mathematics and Econometrics held in Montréal, June 26-30, 2001, under the auspices of CIRANO; particularly the organisers Jérôme Detemple, René Garcia, Eric Renault and Nizar Touzi, for a productive, enjoyable and stimulating week. Thanks are also due to the other participants at the meeting for many illuminating discussions, and for bravely turning up at the end of an already long day of lectures to hear the contents of these talks!

see the main lines of what is going on, and I have to say now that with most of the papers I have read in the literature, I find it easier to take the statement of the problem and work out the dual form for myself than to try to follow the arguments presented in the papers. This is not to say that the paper is redundant, but rather that one can very quickly get to the form of the dual problem, even if proving that the dual and primal problems have equal value remains a substantial task.

There is in fact a unified (and very simple) approach to find the dual form of the problem which works in a wide range of cases[2]. We can think of this as the Pontryagin approach to dynamic programming; or we can think of it in the 'Hamiltonian' language of Bismut (1973), (1975) (see, for example, Malliaris & Brock (1982) for the outline of the method; Chow (1997) also emphasises the efficacy of this approach). To illustrate what I mean, let me now present the method applied to the very simplest example.

Example 0. Suppose we consider an investor who may invest in any of n stocks and in a riskless bank account generating interest at rate r_t. Then the wealth process X of the investor satisfies the dynamics

$$dX_t = r_t X_t dt + \theta_t(\sigma_t dW_t + (b_t - r_t \mathbf{1})dt), \quad X_0 = x, \tag{1.1}$$

where all processes are adapted to the filtration of the standard d-dimensional Brownian motion W, σ takes values in the set of $n \times d$ matrices, and all other processes have the dimensions implied by (1.1)[3]. The process θ is the vector of amounts of wealth invested in each of the stocks. The investor aims to find

$$\sup E\Big[\ U(X_T)\ \Big], \tag{1.2}$$

where $T > 0$ is a fixed time-horizon, and the function $U(\cdot)$ is strictly increasing, strictly concave, and satisfies the Inada conditions[4].

Now we view the dynamics (1.1) of X as some *constraint* to be satisfied by X, and we turn the constrained optimisation problem (1.2) into an *unconstrained* optimisation problem by introducing appropriate Lagrange multipliers. To do this, let the positive process Y satisfy[5]

$$dY_t = Y_t(\beta_t dW_t + \alpha_t dt), \tag{1.3}$$

[2]... but see Section 6.

[3]So, for example, $\mathbf{1}$ is the column n-vector all of whose entries are 1.

[4]$\lim_{x\downarrow 0} U'(x) = \infty$, $\lim_{x\uparrow\infty} U'(x) = 0$. For concreteness, we are assuming that X must remain non-negative.

[5]It is not necessary to express Y in exponential form, but it turns out to be more convenient to do this in any example where processes are constrained to be non-negative.

and consider the integral $\int_0^T Y_s dX_s$. On the one hand, integration by parts gives

$$\int_0^T Y_s dX_s = X_T Y_T - X_0 Y_0 - \int_0^T X_s dY_s - [X,Y]_T, \tag{1.4}$$

and on the other we have (provided constraint/dynamic (1.1) holds)

$$\int_0^T Y_s dX_s = \int_0^T Y_s \theta_s \sigma_s dW_s + \int_0^T Y_s \{ r_s X_s + \theta_s (b_s - r_s \mathbf{1}) \} ds. \tag{1.5}$$

Assuming that expectations of stochastic integrals with respect to W vanish, the expectation of $\int_0^T Y_s dX_s$ is from (1.4)

$$E\Big[X_T Y_T - X_0 Y_0 - \int_0^T Y_s \{ \alpha_s X_s + \theta_s \sigma_s \beta_s \} ds \Big], \tag{1.6}$$

and from (1.5)

$$E\Big[\int_0^T Y_s \{ r_s X_s + \theta_s (b_s - r_s \mathbf{1}) \} ds \Big]. \tag{1.7}$$

Since these two expressions must be equal for any feasible X, we have that the Lagrangian

$$\begin{aligned} \Lambda(Y) &\equiv \sup_{X \geq 0, \theta} E\Big[U(X_T) + \int_0^T Y_s \{ r_s X_s + \theta_s (b_s - r_s \mathbf{1}) \} ds \\ &\qquad - X_T Y_T + X_0 Y_0 + \int_0^T Y_s \{ \alpha_s X_s + \theta_s \sigma_s \beta_s \} ds \Big] \\ &= \sup_{X \geq 0, \theta} E\Big[U(X_T) - X_T Y_T + X_0 Y_0 \\ &\qquad + \int_0^T Y_s \{ r_s X_s + \theta_s (b_s - r_s \mathbf{1}) + \alpha_s X_s + \theta_s \sigma_s \beta_s \} ds \Big] \end{aligned} \tag{1.8}$$

is an upper bound for the value (1.2) whatever Y we take, and will hopefully be equal to it if we minimise over Y.

Now the maximisation of (1.8) over $X_T \geq 0$ is very easy; we obtain

$$\begin{aligned} \Lambda(Y) &= \sup_{X \geq 0, \theta} E\Big[\tilde{U}(Y_T) + X_0 Y_0 \\ &\qquad + \int_0^T Y_s \{ r_s X_s + \theta_s (b_s - r_s \mathbf{1}) + \alpha_s X_s + \theta_s \sigma_s \beta_s \} ds \Big], \end{aligned}$$

where $\tilde{U}(y) \equiv \sup_x [U(x) - xy]$ is the convex dual of U. The maximisation over $X_s \geq 0$ results in a finite value if and only if the complementary slackness condition

$$r_s + \alpha_s \leq 0 \tag{1.9}$$

holds, and maximisation over θ_s results in a finite value if and only if the complementary slackness condition

$$\sigma_s \beta_s + b_s - r_s \mathbf{1} = 0 \tag{1.10}$$

holds. The maximised value is then

$$\Lambda(Y) = E\Big[\tilde{U}(Y_T) + X_0 Y_0\Big]. \tag{1.11}$$

The dual problem therefore ought to be

$$\inf_Y \Lambda(Y) = \inf_Y E\Big[\tilde{U}(Y_T) + X_0 Y_0\Big], \tag{1.12}$$

where Y is a positive process given by (1.3), where α and β are understood to satisfy the complementary slackness conditions (1.9) and (1.10). In fact, since the dual function $\tilde{U}(\cdot)$ is decreasing, a little thought shows that we want Y to be big, so that the 'discount rate' α will be as large as it can be, that is, the inequality (1.9) will actually hold with equality.

We can interpret the multiplier process Y, now written as

$$Y_t = Y_0 \exp\{-\int_0^t r_s ds\}.Z_t,$$

as the product of the initial value Y_0, the riskless discounting term $\exp(-\int_0^t r_s ds)$, and a (change-of-measure) martingale Z, whose effect is to convert the rates of return of all of the stocks into the riskless rate. In the case where $n = d$ and σ has bounded inverse, we find the familiar result of Karatzas, Lehoczky & Shreve (1987), for example, that the marginal utility of optimal wealth is the pricing kernel, or state-price density.

The informal argument just given leads quickly to a candidate for the dual problem; to summarise, the key elements of the approach are:

(a) *write down the dynamics;*

(b) *introduce a Lagrangian semimartingale Y, often in exponential form;*

(c) *transform the dynamics using integration-by-parts;*

(d) *assemble the Lagrangian, and by inspection find the maximum, along with any dual feasibility and complementary slackness conditions.*

We shall see this approach used repeatedly through these lectures; it is a mechanistic way of discovering the dual problem of any given primal problem.

How close to a proof is the argument just given? At first sight, there seem to be big gaps, particularly in the assumption that means of stochastic integrals with respect to local martingales should be zero. But on the other hand, we are looking at a Lagrangian problem, and provided we can *guess* the optimal solution, we we should be able then to *verify* it using little more than the fact that a concave function is bounded above by any supporting hyperplane. So the argument would go that if we have optimal X^*, we would define the dual variable $Y_T^* = U'(X_T^*)$, and simply confirm that the Lagrangian with this choice of Y is maximised at X^*. However, there are problems with this; firstly, we do not know that the supremum is a maximum, and we may have to build X^* as a limit (in what topology?) of approximating X^n; secondly, as we shall soon see in a more general example, the marginal utility of optimal wealth is not necessarily a state-price density. The simple heuristic given above comes tantalisingly close to being a proof of what we want; we can see it, but we are in fact still separated from it by a deep chasm. Getting across this still requires significant effort, though later in Section 3 we shall build a bridge to allow us to cross the chasm - though even this may not be easy to cross.

2 Dual problems made concrete.

Here are some further examples to get practice on.

Example 1. The investor of Example 0 now consumes from his wealth at rate c_t at time t, so that the dynamics of his wealth process becomes

$$dX_t = r_t X_t dt + \theta_t(\sigma_t dW_t + (b_t - r_t \mathbf{1})dt) - c_t dt, \quad X_0 = x, \tag{2.1}$$

His objective now is to find

$$\sup E\Big[\int_0^T U(s, c_s) ds + U(T, X_T) \Big], \tag{2.2}$$

where $T > 0$ is a fixed time-horizon, and for each $0 \le s \le T$ the function $U(s, \cdot)$ is strictly increasing, strictly concave, and satisfies the Inada conditions.

Exercise 1. Apply the general approach given above to show that the dual problem is to find

$$\inf_Y \Lambda(Y) \equiv \inf_Y E\Big[\int_0^T \tilde{U}(s, Y_s) ds + \tilde{U}(T, Y_T) + X_0 Y_0\Big], \tag{2.3}$$

where Y is a positive process satisfying

$$dY_t = Y_t(\beta_t dW_t + \alpha_t dt), \tag{2.4}$$

with

$$\alpha_t = -r_t, \quad \sigma_s \beta_s + b_s - r_s \mathbf{1} = 0 \tag{2.5}$$

Example 2. (El Karoui & Quenez (1995)). In this problem, the agent's wealth once again obeys the dynamics (2.1), but now the objective is to find the *super-replication price*, that is, the smallest value of x such that by judicious choice of θ and c he can ensure that

$$X_T \geq B \quad \text{a.s.},$$

where B is some given $\mathcal{F}_T$-measurable random variable.

EXERCISE 2. Replacing the objective (1.2) from Example 0 by

$$\sup E[u_0(X_T - B)],$$

where $u_0(x) = -\infty$ if $x < 0$, $u_0(x) = 0$ if $x \geq 0$, show that the super-replication price is

$$\sup_{\beta \in \mathcal{B}_0} E[BY_T(\beta)], \tag{2.6}$$

where $Y_t(\beta)$ is the solution to

$$dY_t = Y_t(-r_t dt + \beta_t dW_t), \quad Y_0 = 1,$$

and $\mathcal{B}_0 \equiv \{\text{adapted } \beta \text{ such that } \sigma_s \beta_s + b_s - r_s \equiv 0\}$.

REMARKS. This example shows that we need in general to allow the utility to depend on ω. The super-replication price (2.6) can equally be expressed as

$$\sup_{Q \in \mathcal{M}} E^Q[\,\exp(-\int_0^T r_s ds)B],$$

where $\mathcal{M}$ denotes the set of equivalent martingale measures.

Example 3. (Kramkov & Schachermayer (1999)). This example is the general form of Example 0. In this situation, the asset price processes S are general non-negative semimartingales, and the attainable wealths are random variables X_T which can be expressed in the form

$$X_T = x + \int_0^T H_u dS_u$$

for some previsible H such that the process $X_t \equiv x + \int_0^t H_u dS_u$ remains non-negative.

EXERCISE 3A. If $\mathcal{X}(x)$ denotes the set of such random variables X_T, and if $\mathcal{Y}$ denotes the set of all positive processes Y such that $Y_t X_t$ is a supermartingale for all $X \in \mathcal{X}(x)$, show that the dual form of the problem is

$$\sup_{X_T \in \mathcal{X}(x)} E[U(X_T)] = \inf_{Y \in \mathcal{Y}} E[\tilde{U}(Y_T) + xY_0]. \tag{2.7}$$

EXERCISE 3B. (based on Example 5.1 bis of Kramkov & Schachermayer (1999)). Consider a two-period model where the price process is (S_0, S_1), with $S_0 \equiv 1$, and S_1 taking one of a sequence $(x_n)_{n\geq 0}$ of values decreasing to zero, with probabilities $(p_n)_{n\geq 0}$. Suppose that $x_0 = 2$, $x_1 = 1$, and suppose that

$$\frac{p_0}{\sqrt{2}} > \sum_{n\geq 1} \frac{p_n(1-x_n)}{\sqrt{x_n}}. \tag{2.8}$$

The agent has utility $U(x) = \sqrt{x}$, initial wealth 1, and his portfolio choice consists simply of choosing the number $\lambda \in [-1, 1]$ of shares to be held. If he holds λ, his expected utility is

$$EU(X_1) = \sum_{n\geq 0} p_n \sqrt{1 - \lambda + \lambda x_n}. \tag{2.9}$$

Prove that his optimal choice is $\lambda = 1$, but that $U'(X_1^*)$ is not in general a (multiple of an) equivalent martingale measure:

$$E[U'(X_1^*)(S_1 - S_0)] \neq 0. \tag{2.10}$$

Example 4. (Cuoco & Liu (2000)). This is an important example, generalising a number of other papers in the subject: Cvitanić & Karatzas (1992, 1993), El Karoui, Peng & Quenez (1997), Cuoco & Cvitanić (1998), for example. The wealth process X of the agent satisfies

$$dX_t = X_t \Big[r_t dt + \pi_t \cdot \{\sigma_t dW_t + (b_t - r_t \mathbf{1})dt\} + g(t, \pi_t)dt \Big] - c_t dt, \qquad X_0 = x, \tag{2.11}$$

where W is an n-dimensional Brownian motion, b, r, $V \equiv \sigma\sigma^T$, V^{-1} are all bounded processes, and there is a uniform Lipschitz bound on g: for some $\gamma < \infty$,

$$|g(t, x, \omega) - g(t, y, \omega)| \leq \gamma |x - y|$$

for all x, y, t and ω.

The only unconventional term in the dynamics (2.11) is the term involving g, about which we assume:

(i) for each $x \in \mathbb{R}^n$, $(t, \omega) \mapsto g(t, x, \omega)$ is an optional process;

(ii) for each $t \in [0,T]$ and $\omega \in \Omega$, $x \mapsto g(t,x,\omega)$ is concave and upper semi-continuous.

(iii) $g(t,0,\omega) = 0$ for all $t \in [0,T]$ and $\omega \in \Omega$.

The agent has the objective of maximising

$$E\Big[\int_0^T U(s,c_s)ds + U(T,X_T) \Big], \tag{2.12}$$

where we assume that for every $t \in [0,T]$ the map $c \mapsto U(t,c)$ is strictly increasing, strictly concave, and satisfies the Inada conditions.

Exercise 4. Show that the dual problem is to find

$$\inf_Y E\Big[\int_0^T \tilde{U}(t,Y_t)\,dt + \tilde{U}(T,Y_T) + xY_0 \Big] \tag{2.13}$$

where the process Y solves

$$Y_t^{-1}dY_t = V_t^{-1}(r_t\mathbf{1} - b_t - \nu_t) \cdot \sigma_t dW_t - (r_t + \tilde{g}(t,\nu_t))dt \tag{2.14}$$

for some adapted process ν bounded by γ, and where $\tilde{g}$ is the convex dual of g.

Example 5. (Cvitanic & Karatzas (1996)). This is a simple example incorporating transaction costs, where the holdings X_t of cash and the holdings Y_t of the sole share at time t obey

$$dX_t = r_tX_tdt + (1-\varepsilon)dM_t - (1+\delta)dL_t - c_tdt, \tag{2.15}$$

$$dY_t = Y_t(\sigma_t dW_t + \rho_t dt) - dM_t + dL_t, \tag{2.16}$$

where M and L are increasing processes, with the usual uniform boundedness assumptions on σ, ρ, σ^{-1} and r. The investor starts with initial holdings $(X_0,Y_0) = (x,y)$, and chooses the pair (L,M) and the consumption rate c so as to achieve his objective. Suppose that this is to

$$\sup E\Big[\int_0^T U(c_s)ds + u(X_T,Y_T)\Big],$$

where we restrict to strategies lying always in the solvency region:

$$X_t + (1-\varepsilon)Y_t \geq 0, \quad X_t + (1+\delta)Y_t \geq 0 \quad \forall t$$

Exercise 5. By introducing Lagrangian semimartingales

$$d\xi_t = \xi_t(\alpha_t dW_t + \beta_t dt),$$

$$d\eta_t = \eta_t(a_t dW_t + b_t dt),$$

show that the dual form of the problem is

$$\inf E\Big[\int_0^T \tilde{U}(t,\xi_t)dt + \tilde{u}(\xi_T,\eta_T) + x\xi_0 + y\eta_0\Big],$$

where the dual feasibility conditions to be satisfied are

$$\begin{aligned}\beta_t &= -r_t \\ b_t &= -\rho_t - \sigma_t a_t \\ 1-\varepsilon \leq & \tfrac{\eta_t}{\xi_t} \leq 1+\delta,\end{aligned}$$

and $\tilde{U}$, $\tilde{u}$ are the convex dual functions of U, u respectively.

Remarks. The Lagrangian semimartingales of this example are related to the dual processes (Z^0, Z^1) of Cvitanic & Karatzas by

$$Z_t^0 = B_t\xi_t, \quad Z_t^1 = S_t\eta_t,$$

where B is the bond price process, and S is the stock price process solving

$$dB_t = B_t dt, \quad dS_t = S_t(\sigma_t dW_t + \rho_t dt).$$

Example 6. (Broadie, Cvitanic & Soner (1998)). This interesting example finds the minimum super-replicating price of a European-style contingent claim with a constraint on the portfolio process. Thus we are looking at the problem of Example 2, with the dynamics (2.11) of Example 4, specialized by assuming that σ, b and r are positive constants, and that g takes the form

$$g(t,x) = 0 \quad \text{if } x \in C; \quad = -\infty \quad \text{if } x \notin C,$$

where C is some closed convex set. We suppose that the contingent claim to be super-replicated, B, takes the form $B = \varphi(S)$ for some non-negative lower semi-continuous function φ, where S is the vector of share prices, solving

$$dS_t^i = S_t^i\Big[\sum_j \sigma_{ij} dW_t^j + \rho_i dt\Big].$$

The interest rate r, volatility matrix σ and drift ρ are all assumed constant, and σ is square and invertible.

Exercise 6. (Cvitanic & Karatzas (1993).) Show that the super-replication price for B is given by

$$\sup_\nu E[Y_T(\nu)B],$$

where $Y(\nu)$ solves (2.14) with initial condition $Y_0 = 1$.

3 Dual problems made difficult.

Example 5 shows us that a formulation broad enough to embrace problems with transaction costs has to consider vector-valued asset processes; it is not sufficient to consider the aggregate wealth of the investor. This can be done, and *is* done in Klein & Rogers (2003), but in the present context we shall restrict ourselves to a univariate formulation of the problem. This saves us from a certain amount of careful convex analysis, which is not particularly difficult, and gives a result which will cover all the earlier examples apart from Example 5.

The main result below, Theorem 1, proves that under certain conditions, the value of the primal problem, expressed as a supremum over some set, is equal to the value of the dual problem, expressed as an infimum over some other set. It is important to emphasise that the Theorem does *not* say that the supremum in the primal problem is attained in the set, because such a result is not true in general without further conditions, and is typically very deep: see the paper of Kramkov & Schachermayer (1999), which proves that in the situation of Example 3 a further condition on the utility is needed in general to deduce that the value of the primal problem is attained. The result presented here is at its heart an application of the Minimax Theorem, and the argument is modelled on the argument of Kramkov & Schachermayer (1999).

To state the result, we set up some notation and introduce various conditions, a few of which (labelled in bold face) are typically the most difficult to check. Let $(S, \mathcal{S}, \mu)$ be some finite measure space, and let $L^0_+(S, \mathcal{S}, \mu)$ denote the cone of non-negative functions in $L^0(S, \mathcal{S}, \mu)$, a closed convex set usually abbreviated to L^0_+. We shall suppose that for each $x \geq 0$ we have a subset $\mathcal{X}(x)$ of L^0_+ with the properties

(X1) *$\mathcal{X}(x)$ is convex;*

(X2) *$\mathcal{X}(\lambda x) = \lambda \mathcal{X}(x)$ for all $\lambda > 0$;*

(X3) *if $g \in L^0_+$ and $g \leq f$ for some $f \in \mathcal{X}(x)$, then $g \in \mathcal{X}(x)$ also;*

(X4) *the constant function $\mathbf{1} : s \mapsto 1$ is in $\mathcal{X}$,*

where we have used the notation

$$\mathcal{X} \equiv \bigcup_{x \geq 0} \mathcal{X}(x) = \bigcup_{x \geq 0} x\mathcal{X}(1) \tag{3.1}$$

in stating (X4).

For the dual part of the story, we need for each $y \geq 0$ a subset $\mathcal{Y}(y) \subseteq L^0_+$ with the property

(Y1) $\mathcal{Y}(y)$ *is convex;*

(Y2) *for each $y \geq 0$, the set $\mathcal{Y}(y)$ is closed under convergence in μ-measure.*

We introduce the notation

$$\mathcal{Y} \equiv \bigcup_{y \geq 0} \mathcal{Y}(y) \tag{3.2}$$

for future use.

The primal and dual quantities are related by the key polarity property

(XY) *for all $f \in \mathcal{X}$ and $y \geq 0$*

$$\sup_{g \in \mathcal{Y}(y)} \int fg \, d\mu = \inf_{x \in \Psi(f)} xy$$

where we have used the notation

$$\Psi(f) = \{x \geq 0 : f \in \mathcal{X}(x)\}.$$

Properties (X2) and (X3) give us immediately that for any $f \in \mathcal{X}$ there is some $\xi(f) \geq 0$ such that

$$\Psi(f) = (\xi(f), \infty) \;\text{ or }\; [\xi(f), \infty);$$

as yet, we do not know whether the lower bound is in $\Psi(f)$ or not, but we *can* say for $f \in \mathcal{X}(x)$ we must have $\xi(f) \leq x$. It also follows from (XY) that

$$\int fg \, d\mu, \leq xy \qquad f \in \mathcal{X}(x), g \in \mathcal{Y}(y). \tag{3.3}$$

Using (X4), we see from (3.3) that in fact $\mathcal{Y} \subseteq L^1_+$.

Important remark. We shall see in examples that often we take in (Y1) some convex set $\mathcal{Y}_0(y)$ of exponential semimartingales started from y, and it is in general not at all clear that the condition (Y2) will be satisfied for these. However, if we let $\mathcal{Y}(y)$ denote the closure in $L^0(\mu)$ of $\mathcal{Y}_0(y)$, this remains convex, now satisfies (Y2) by definition, and by Fatou's lemma

$$\sup_{g \in \mathcal{Y}(y)} \int fg d\mu = \sup_{g \in \mathcal{Y}_0(y)} \int fg d\mu,$$

so all we need to confirm (XY) is to check the statements for $g \in \mathcal{Y}_0(y)$. There is of course a price to pay, and that is that the statement of the main result is somewhat weaker.

Finally, we shall need a utility function $U : S \times \mathbb{R}^+ \to \mathbb{R} \cup \{-\infty\}$ with the basic properties

(U1) *$s \mapsto U(s,x)$ is $\mathcal{S}$-measurable for all $x \geq 0$;*

(U2) *$x \mapsto U(s,x)$ is concave, differentiable, strictly increasing, and finite-valued on $(0,\infty)$ for every $s \in \mathcal{S}$.*

We shall without comment assume that the definition of U has been extended to the whole of $S \times \mathbb{R}$ by setting $U(s,x) = -\infty$ if $x < 0$. Differentiability is not essential, but makes some subsequent statements easier.

We also impose the Inada-type conditions:

(U3) *if*

$$\varepsilon_n(s) \equiv U'(s,n), \tag{3.4}$$

we suppose that

$$\varepsilon_n(s) \to 0 \quad \mu - \text{a.e.} \tag{3.5}$$

as $n \to \infty$, and that there exists some n_0 such that

$$\int |\varepsilon_{n_0}(s)| \, \mu(ds) < \infty. \tag{3.6}$$

One consequence of this is that

$$U(s,x)/x \to 0 \quad (x \to \infty), \tag{3.7}$$

and another is that for any $z > 0$, the supremum defining the convex dual function $\tilde{U}$ is attained:

$$\begin{aligned} \tilde{U}(s,z) &\equiv \sup_{x>0}\{U(s,x) - xz\} \\ &= \max_{x>0}\{U(s,x) - xz\}. \end{aligned} \tag{3.8}$$

(U4) *the concave function*

$$\underline{u}(\lambda) \equiv \inf_{s \in \mathcal{S}} U(s,\lambda)$$

is finite-valued on $(0,\infty)$ and satisfies the Inada condition

$$\lim_{\lambda \downarrow 0} \frac{\partial \underline{u}}{\partial \lambda} = \infty;$$

Important remark[6] . We can in fact relax the condition (U4) to the simpler

[6]Thanks to Nizar Touzi for noticing that the Inada condition at zero is unnecessary.

(U4') *the concave function*

$$\underline{u}(\lambda) \equiv \inf_{s \in \mathcal{S}} U(s, \lambda)$$

is finite-valued on $(0, \infty)$.

The reason, explained in more detail in Section 6, is that we can always approximate a given utility uniformly to within any given $\varepsilon > 0$ by one satisfying the Inada condition at 0.

We impose one last (very slight) condition of a technical nature:

(U5) *there exists* $\psi \in \mathcal{X}$, *strictly positive, such that for all* $\varepsilon \in (0, 1)$

$$U'(s, \varepsilon\psi(s)) \in L^1(S, \mathcal{S}, \mu);$$

Next we define the functions $u : \mathbb{R}^+ \to [-\infty, \infty)$ and $\tilde{u} : \mathbb{R}^+ \to (-\infty, \infty]$ by

$$u(x) \equiv \sup_{f \in \mathcal{X}(x)} \int U(s, f(s))\mu(ds) \tag{3.9}$$

and

$$\tilde{u}(y) \equiv \inf_{g \in \mathcal{Y}(y)} \int \tilde{U}(s, g(s))\mu(ds). \tag{3.10}$$

To avoid vacuous statements, we make the following finiteness assumption:

(F) for some $f_0 \in \mathcal{X}$ and $g_0 \in \mathcal{Y}$ we have

$$\begin{aligned} \int U(s, f_0(s))\mu(ds) &> -\infty, \\ \int \tilde{U}(s, g_0(s))\mu(ds) &< \infty. \end{aligned}$$

Notice immediately one simple consequence of (F) and (3.3): if $f \in \mathcal{X}(x)$ and $g \in \mathcal{Y}(y)$,

$$\begin{aligned} \int U(s, f(s))\, \mu(ds) &\leq \int \Big[\, U(s, f(s)) - f(s)g(s)\, \Big]\, \mu(ds) + xy \\ &\leq \int \tilde{U}(s, g(s))\, \mu(ds) + xy. \end{aligned} \tag{3.11}$$

Taking $g = g_0$ in this inequality tells us that u is finite-valued, and taking $f = f_0$ tells us that $\tilde{u}$ is finite-valued.

Theorem 1 *Under the conditions stated above, the functions u and $\tilde{u}$ are dual:*

$$\tilde{u}(y) = \sup_{x \geq 0}[\, u(x) - xy \,], \tag{3.12}$$

$$u(x) = \inf_{y \geq 0}[\, \tilde{u}(y) + xy \,]. \tag{3.13}$$

Proof. Firstly, notice that part of what we have to prove is very easy: indeed, using the inequality (3.11), by taking the supremum over $f \in \mathcal{X}(x)$ and the infimum over $g \in \mathcal{Y}(y)$ we have that

$$\tilde{u}(y) \geq u(x) - xy \tag{3.14}$$

for any non-negative x and y. The other inequality is considerably more difficult, and is an application of the Minimax Theorem.

Define the function $\Phi : \mathcal{X} \times \mathcal{Y} \to [-\infty, \infty)$ by

$$\Phi(f, g) \equiv \int [\, U(s, f(s)) - f(s)g(s) \,]\, \mu(ds), \tag{3.15}$$

and introduce the sets

$$\mathcal{B}_n \equiv \{ f \in L^{\infty}_{+}(S, \mathcal{S}, \mu) : 0 \leq f(s) \leq n \ \forall s \}. \tag{3.16}$$

Then $\mathcal{B}_n$ is convex, and compact in the topology $\sigma(L^{\infty}, L^1)$. We need the following result.

Lemma 1 *For each $y \geq 0$, for each $g \in \mathcal{Y}(y)$, the map $f \mapsto \Phi(f, g)$ is upper semicontinuous on $\mathcal{B}_n$ and is sup-compact: for all a*

$$\{ f \in \mathcal{B}_n : \Phi(f, g) \geq a \} \quad \text{is } \sigma(L^{\infty}, L^1)\text{-compact.}$$

Proof. The map $f \mapsto \int fg d\mu$ is plainly continuous in $\sigma(L^{\infty}, L^1)$ on $\mathcal{B}_n$, so it is sufficient to prove the upper semicontinuity assertion in the case $g = 0$,

$$f \mapsto \int U(s, f(s))\, \mu(ds).$$

Once we have upper semicontinuity, the compactness statement is obvious. So the task is to prove that for any $a \in \mathbb{R}$, the set

$$\begin{aligned}\{f \in \mathcal{B}_n : \int U(s, f(s))\, \mu(ds) &\geq a\} \\ &= \bigcap_{\varepsilon>0} \{f \in \mathcal{B}_n : \int U(s, f(s) + \varepsilon\psi(s))\, \mu(ds) \geq a\}\end{aligned}$$

is $\sigma(L^\infty, L^1)$-closed. The equality of these two sets is immediate from the Monotone Convergence Theorem, the fact that $\psi \in \mathcal{X}$, and the fact that $U(s, \cdot)$ is increasing for all s. We shall prove that for each $\varepsilon > 0$ the set

$$N_\varepsilon = \{f \in \mathcal{B}_n : \int U(s, f(s) + \varepsilon\psi(s))\, \mu(ds) < a\}$$

is open in $\sigma(L^\infty, L^1)$. Indeed, if $h \in \mathcal{B}_n$ is such that

$$\int U(s, h(s) + \varepsilon\psi(s))\, \mu(ds) = a - \delta < a,$$

we have by (U2) that for any $f \in \mathcal{B}_n$

$$\begin{aligned}&\int U(s, f(s) + \varepsilon\psi(s))\, \mu(ds) \\ &\leq \int \Big[U(s, h(s) + \varepsilon\psi(s)) + (f(s) - h(s))U'(s, h(s) + \varepsilon\psi(s))\Big]\, \mu(ds) \\ &\leq a - \delta + \int (f(s) - h(s))U'(s, h(s) + \varepsilon\psi(s))\, \mu(ds)\end{aligned}$$

Since $U'(s, h(s) + \varepsilon\psi(s)) \in L^1(S, \mathcal{S}, \mu)$ by (U5), this exhibits a $\sigma(L^\infty, L^1)$-open neighbourhood of h which is contained in N_ε, as required. $\square$

We now need the Minimax Theorem, Theorem 7 on p 319 of Aubin & Ekeland (1984), which we state here for completeness, expressed in notation adapted to the current context.

Minimax Theorem. *Let B and Y be convex subsets of vector spaces, B being equipped with a topology. If*

(MM1) for all $g \in Y$, $f \mapsto \Phi(f, g)$ is concave and upper semicontinuous;

(MM2) for some $g_0 \in Y$, $f \mapsto \Phi(f, g_0)$ is sup-compact;

(MM3) for all $f \in B$, $g \mapsto \Phi(f, g)$ is convex,

then

$$\sup_{f \in B} \inf_{g \in Y} \Phi(f,g) = \inf_{g \in Y} \sup_{f \in B} \Phi(f,g),$$

and the supremum on the left-hand side is attained at some $\bar{f} \in B$.

We therefore have

$$\sup_{f \in \mathcal{B}_n} \inf_{g \in \mathcal{Y}(y)} \Phi(f,g) = \inf_{g \in \mathcal{Y}(y)} \sup_{f \in \mathcal{B}_n} \Phi(f,g). \tag{3.17}$$

From this,

$$\sup_{f \in \mathcal{B}_n} \inf_{g \in \mathcal{Y}(y)} \Phi(f,g) = \inf_{g \in \mathcal{Y}(y)} \int \tilde{U}_n(s, g(s))\, \mu(ds) \equiv \tilde{u}_n(y), \tag{3.18}$$

say, where

$$\tilde{U}_n(s,z) \equiv \sup\{U(s,x) - zx : 0 \le x \le n\} \uparrow \tilde{U}(s,z). \tag{3.19}$$

Consequently, $\tilde{u}_n(y) \le \tilde{u}(y)$.

Using the property (XY) going from the second to the third line, we estimate

$$\begin{aligned}
\tilde{u}_n(y) = \sup_{f \in \mathcal{B}_n} \inf_{g \in \mathcal{Y}(y)} \Phi(f,g) &= \sup_{f \in \mathcal{B}_n} \inf_{g \in \mathcal{Y}(y)} \int \{U(s,f(s)) - f(s)g(s)\}\mu(ds) \\
&= \sup_{f \in \mathcal{B}_n} \Big[\int U(s,f(s))\mu(ds) - \sup_{g \in \mathcal{Y}(y)} \int fg\, d\mu\Big] \\
&= \sup_{f \in \mathcal{B}_n} \Big[\int U(s,f(s))\mu(ds) - \inf_{x \in \Psi(f)} xy\Big] \\
&= \sup_{f \in \mathcal{B}_n} \sup_{x \in \Psi(f)} \Big[\int U(s,f(s))\mu(ds) - xy\Big] \\
&\le \sup_{f \in \mathcal{X}} \sup_{x \in \Psi(f)} \Big[\int U(s,f(s))\mu(ds) - xy\Big] \\
&= \sup_{x \in C} \sup_{f \in \mathcal{X}(x)} \Big[\int U(s,f(s))\mu(ds) - xy\Big] \\
&= \sup_{x \in C} \Big[u(x) - xy\Big] \\
&\le \tilde{u}(y)
\end{aligned}$$

The $\tilde{u}_n(y)$ clearly increase with n, so the proof will be complete provided we can prove that

$$\lim_n \tilde{u}_n(y) = \tilde{u}(y), \tag{3.20}$$

Suppose that $g_n \in \mathcal{Y}(y)$ are such that

$$\tilde{u}_n(y) \leq \int \tilde{U}_n(s, g_n(s))\mu(ds) \leq \tilde{u}_n(y) + n^{-1}. \tag{3.21}$$

Using Lemma A1.1 of Delbaen & Schachermayer (1994), we can find a sequence $h_n \in \mathrm{conv}(g_n, g_{n+1}, \ldots)$ in $\mathcal{Y}(y)$ which converge μ-almost everywhere to a function h taking values in $[0, \infty]$; because of (Y2), $h \in \mathcal{Y}(y)$. Moreover, because of (X4) and (3.3), the limit must be almost everywhere finite, and hence

$$\lim_n \int \tilde{U}_n(s, h_n(s))\mu(ds) = \lim_n \tilde{u}_n(y).$$

From the definition (3.4) of $\varepsilon_n(s)$, it is immediate that

$$\tilde{U}_n(s, z) = \tilde{U}(s, z) \quad \text{if } z \geq \varepsilon_n(s).$$

One last fact is needed, which we prove later.

Proposition 1 *The family* $\{\tilde{U}(s, \varepsilon_n(s) + g(s))^- : g \in \mathcal{Y}(y), n \geq n_0\}$ *is uniformly integrable.*

Using these facts, we have the inequalities

$$\begin{aligned}
\tilde{u}(y) &\leq \int \tilde{U}(s, h(s))\, \mu(ds) \\
&\leq \liminf_n \int \tilde{U}(s, \varepsilon_n(s) + h(s))\, \mu(ds) \\
&\leq \liminf_n \liminf_{m \geq n} \int \tilde{U}(s, \varepsilon_n(s) + h_m(s))\, \mu(ds) \\
&\leq \liminf_n \liminf_{m \geq n} \int \tilde{U}_m(s, \varepsilon_n(s) + h_m(s))\, \mu(ds) \\
&\leq \liminf_n \liminf_{m \geq n} \int \tilde{U}_m(s, h_m(s))\, \mu(ds) \\
&= \lim_n \tilde{u}_n(y) \\
&\leq \tilde{u}(y),
\end{aligned}$$

as required. □

Proof of Proposition 1. The argument here is a slight modification of that of Kramkov & Schachermayer (1999); we include it for completeness. Firstly, we note that

$$\begin{aligned}
-\tilde{U}(s, z) &\equiv \inf_{x \geq 0}\{\, xz - U(s, x) \,\} \\
&\leq \inf_{x \geq 0}\{\, xz - \underline{u}(x) \,\} \\
&\equiv \psi(z),
\end{aligned}$$

say. We suppose that $\sup_a \psi(a) = \infty$, otherwise there is nothing to prove, and let $\varphi : (\psi(0), \infty) \to (0, \infty)$ denote its convex increasing inverse. We have

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = \lim_{y\to\infty} \frac{y}{\psi(y)} = \lim_{t\downarrow 0} \frac{\underline{u}'(t)}{t\underline{u}'(t) - \underline{u}(t)} = \lim_{t\downarrow 0} \frac{\int_t^1 \underline{u}''(ds)}{\int_t^1 s\underline{u}''(ds)} = \infty,$$

using the property (U4). Now we estimate

$$\begin{aligned}\int \varphi(\tilde{U}(s, \varepsilon_n(s) + g(s))^-)\mu(ds) &\le \int \varphi(\max\{0, \psi(\varepsilon_n(s) + g(s))\}\mu(ds) \\ &\le \varphi(0)\mu(S) + \int \varphi(\psi(\varepsilon_n(s) + g(s)))\mu(ds) \\ &= \varphi(0)\mu(S) + \int \{\varepsilon_n(s) + g(s)\}\mu(ds).\end{aligned}$$

This is bounded uniformly in $g \in \mathcal{Y}(y)$, by (X4), (3.3) and (U3). □

There is a useful little corollary of this proposition.

Corollary 1 *For each $y \ge 0$, there is some $g \in \mathcal{Y}(y)$ for which the infimum defining $\tilde{u}(y)$ in (3.10) is attained.*

Differentiability of U implies strict convexity of $\tilde{U}$, which in turn implies uniqueness of the minimising g.

PROOF. Take $g_n \in \mathcal{Y}(y)$ such that

$$\tilde{u}(y) \le \int \tilde{U}(s, g_n(s))\mu(ds) \le \tilde{u}(y) + n^{-1}. \tag{3.22}$$

By again using Lemma A1.1 of Delbaen & Schachermayer (1994) we may suppose that the g_n are μ-almost everywhere convergent to limit g, still satisfying the inequalities (3.22). Now by Proposition 1 and Fatou's lemma,

$$\tilde{u}(y) \le \int \tilde{U}(s, g(s))\mu(ds) \le \liminf_n \int \tilde{U}(s, g_n(s))\mu(ds) \le \tilde{u}(y),$$

as required. The uniqueness assertion is immediate. □

4 Dual problems made honest

So far, I have made a number of soft comments about how to identify the dual problem in a typical situation, and have stated and proved a general abstract result which I claim turns the heuristic recipe into a proved result. All we have to do in any given example is simply to verify the conditions of the previous Section, under which Theorem 1 was proved; in this Section, we will see this verification done in full for the Cuoco-Liu example, and you will be able to judge for yourself just how 'simply' this verification can be done in practice!

To recall the setting, the dynamics of the agent's wealth is given by (2.11):

$$dX_t = X_t \Big[r_t dt + \pi_t \cdot \{\sigma_t dW_t + (b_t - r_t \mathbf{1})dt\} + g(t, \pi_t)dt \Big] - c_t dt, \qquad X_0 = x, \tag{4.1}$$

where the processes b, r, $V \equiv \sigma\sigma^T$, V^{-1} are all bounded processes, and there is a uniform Lipschitz bound on g: for some $\gamma < \infty$,

$$|g(t, x, \omega) - g(t, y, \omega)| \leq \gamma |x - y|$$

for all x, y, t and ω. The function g is assumed to be concave and vanishing at zero in its second argument, and the agent aims to maximise his objective (2.12):

$$E\Big[\int_0^T U(s, c_s)ds + U(T, X_T) \Big], \tag{4.2}$$

where for every $t \in [0, T]$ the map $c \mapsto U(t, c)$ is strictly increasing, strictly concave, and satisfies the Inada conditions.

We believe that the dual form of the problem is given by the solution to Exercise 4, namely, to find

$$\inf_Y E\Big[\int_0^T \tilde{U}(t, Y_t)\, dt + \tilde{U}(T, Y_T) + xY_0 \Big] \tag{4.3}$$

where the process Y solves

$$Y_t^{-1} dY_t = V_t^{-1}(r_t \mathbf{1} - b_t - \nu_t) \cdot \sigma_t dW_t - (r_t + \tilde{g}(t, \nu_t))dt \tag{4.4}$$

for some adapted process ν bounded by γ, and where $\tilde{g}$ is the convex dual of g.

Now we have to cast the problem into the form of Section 3 so that we may apply the main result, Theorem 1. For the finite measure space $(S, \mathcal{S}, \mu)$ we take

$$S = [0, T] \times \Omega, \quad \mathcal{S} = \mathcal{O}[0, T], \quad \mu = (\mathrm{Leb}[0, T] + \delta_T) \times P,$$

where $\mathcal{O}[0, T]$ is the optional[7] σ-field restricted to $[0, T]$. The set $\mathcal{X}(x)$ is the collection of all *bounded* optional $f : S \mapsto \mathbb{R}^+$ such that for some non-negative (X, c) satisfying (4.1), for all ω,

$$f(t, \omega) \leq c(t, \omega), \qquad (0 \leq t < T), \qquad f(T, \omega) \leq X(T, \omega). \tag{4.5}$$

[7]That is, the σ-field generated by the stochastic intervals $[\tau, \infty)$ for all stopping times τ of the Brownian motion. See, for example, Section VI.4 of Rogers & Williams (2000).

REMARK. The assumption that f is bounded is a technical detail without which it appears very hard to prove anything. The conclusion is not in any way weakened by this assumption, though, as we shall discuss at the end.

Next we define $\mathcal{Y}_1(y)$ to be the set of all solutions to (2.14) with initial condition $Y_0 = y$. From this we define the set $\mathcal{Y}_0(y)$ to be the collection of all non-negative adapted processes h such that for some $Y \in \mathcal{Y}_1(y)$

$$h(t,\omega) \leq Y(t,\omega) \quad \mu\text{-almost everywhere.}$$

Finally, we define a utility function $\varphi : S \times \mathbb{R}^+ \mapsto \mathbb{R} \cup \{-\infty\}$ in the obvious way:

$$\varphi((t,\omega),x) = U(t,x),$$

and we shall slightly abuse notation and write U in place of φ henceforth.

We have now defined the objects in terms of which Theorem 1 is stated, and we have to prove that they have the required properties.

(X1) If (X^1, c^1) and (X^2, c^2) solve (4.1) with portfolio processes π^1 and π^2, say, taking any $\theta_1 = 1 - \theta_2 \in [0,1]$ and defining

$$\begin{aligned}
\bar{X} &= \theta^1 X^1 + \theta^2 X^2, \\
\bar{c} &= \theta^1 c^1 + \theta^2 c^2, \\
\bar{\pi} &= \frac{\theta^1 \pi^1 X^1 + \theta^2 \pi^2 X^2}{\bar{X}}
\end{aligned}$$

we find immediately that

$$d\bar{X}_t = \bar{X}_t \Big[\, r_t dt + \bar{\pi}_t \cdot \{\sigma_t dW_t + (b_t - r_t \mathbf{1})dt\} + g(t,\bar{\pi}_t)dt - \Psi_t dt \, \Big] - \bar{c}_t dt,$$

where

$$\Psi_t = g(t,\bar{\pi}_t) - \Big[\theta^1 X_t^1 g(t,\pi_t^1) + \theta^2 X_t^2 g(t,\pi_t^2) \, \Big] / \bar{X} \geq 0,$$

using the concavity of g. It easy to deduce from this that

$$X_t^* - \bar{X}_t \geq 0, \qquad 0 \leq t \leq T,$$

where X^* is the solution to

$$dX_t^* = X_t^* \Big[\, r_t dt + \bar{\pi}_t \cdot \{\sigma_t dW_t + (b_t - r_t \mathbf{1})dt\} + g(t,\bar{\pi}_t)dt \, \Big] - \bar{c}_t dt,$$

starting at x. Hence $(X^*, \bar{c})$ solves (4.1) with portfolio $\bar{\pi}$, and the convex combination $(\bar{X}, \bar{c})$ is in $\mathcal{X}(x)$. Hence $\mathcal{X}(x)$ is convex.

(X2) and (X3) are trivial.

(X4) By taking $\pi = 0$, and using the fact that r is bounded, we see from the dynamics (2.11) that for some small enough $\varepsilon > 0$ we can achieve a constant consumption stream $c_t = \varepsilon$ with terminal wealth $X_T \geq \varepsilon$. This establishes (X4).

(Y1) The proof of convexity of $\mathcal{Y}(y)$ is analogous to the proof of property (X1).

(Y2) Because of the global Lipschitz assumption on g, it can be shown that in fact $\mathcal{Y}_0(y)$ is closed in $L^0(\mu)$ for all $y \geq 0$; see Klein & Rogers (2001) for details. Hereafter, $\mathcal{Y}(y)$ will be defined to be the closure in $L^0(\mu)$ of $\mathcal{Y}_0(y)$; they *are* equal, but no use will be made of this fact.

The properties of the utility, and the finiteness assumption are as quickly dealt with: properties (U1) and (U2) are evident, while the remaining properties must be checked on each particular case. For example, if the utility has separable form

$$U(s, c) = h(s)f(c)$$

then provided h is bounded, and f is strictly increasing, concave and satisfies the Inada conditions, the conditions (U3)–(U5) are satisfied. For the finiteness condition (F), we must once again check this for each particular case.

(XY) So far so good; now the work begins. The heart of the proof lies in establishing the duality relation (XY) but in reality only one half of (XY) presents any challenges.

If (c, X) solves (4.1) and if Y solves (4.3), then using Itô's formula gives us

$$\begin{aligned} d(X_tY_t) &= X_tY_t\{\pi_t \cdot \sigma_t dW_t + (r_t\mathbf{1} - b_t - \nu_t) \cdot \sigma_t^{-1} dW_t \\ &\qquad + (-\pi_t \cdot \nu_t + g(t, \pi_t) - \tilde{g}(t, \nu_t))dt\} - c_tY_tdt \\ &\doteq X_tY_t[(g(t, \pi_t) - \pi_t \cdot \nu_t - \tilde{g}(t, \nu_t)]dt - c_tY_tdt, \end{aligned} \tag{4.6}$$

where the symbol $\doteq$ signifies that the two sides differ by a local martingale. From this, using the definition of $\tilde{g}$, we conclude that

$$X_tY_t + \int_0^t Y_s c_s \, ds \quad \text{is a non-negative supermartingale,}$$

which leads immediately to the inequality

$$X_0Y_0 \geq E[\int_0^T Y_s c_s \, ds + X_TY_T].$$

Thus we have half of (XY); if $f \in \mathcal{X}(x)$, then

$$\sup_{h \in \mathcal{Y}(y)} \int fh \, d\mu \leq xy,$$

and so

$$\sup_{h\in\mathcal{Y}(y)}\int fh\,d\mu \le \inf_{x\in\Psi(f)} xy.$$

REMARK. If we took *any* f dominated as at (4.5), but not necessarily bounded, the above analysis still holds good, and $\int fh\,d\mu \le xy$ for all $h\in\mathcal{Y}(y)$.

What remains now is to prove that if $f\in\mathcal{X}$ and

$$\sup_{h\in\mathcal{Y}(1)}\int fh\,d\mu \equiv \xi \le x \tag{4.7}$$

then $f\in\mathcal{X}(x)$, for which it is evidently equivalent to prove that $f\in\mathcal{X}(\xi)$ in view of (X2). Notice the interpretation of what we are required to do here. It could be that the given $f\in\mathcal{X}$ were dominated as at (4.5) by some (c,X) which came from a very large initial wealth x_0, but that the value ξ were much smaller than x_0. What we now have to do is to show that the consumption plan and terminal wealth defined by f can actually be financed by the smaller initial wealth ξ.

The argument requires three steps:

Step 1: Show that the supremum in (4.7) is attained at some $Y^*\in\mathcal{Y}_1(1)$:

$$\xi = E\Big[\int_0^T Y_s^* f_s\,ds + Y_T^* f_T\Big]; \tag{4.8}$$

Step 2: Use the Y^* from Step 1 to construct a (conventional) market in which the desired consumption stream and terminal wealth f can be achieved by replication, using investment process π^* with initial wealth ξ;

Step 3: By considering the process ν related to π^* by duality, show that in fact the investment process π^* replicates f in the original market.

Here is how the three steps of the argument are carried out.

Step 1. We prove the following.

Proposition 2 *There exists $Y^*\in\mathcal{Y}(1)$, with corresponding process ν^* in (4.4), such that*

$$\xi = E[\int_0^T Y_s^* f_s\,ds + f_T Y_T^*]. \tag{4.9}$$

PROOF. There exist processes $\nu^{(n)}$ bounded by the constant γ such that if $Y^{(n)}$ is the solution to (4.4) using $\nu=\nu^{(n)}$ with initial value 1, then

$$E[\int_0^T Y_s^{(n)} f_s\,ds + f_T Y_T^{(n)}] > \xi - 2^{-n}. \tag{4.10}$$

We need to deduce convergence of the $\nu^{(n)}$ and the $Y^{(n)}$. Using once again Lemma A1.1 of Delbaen & Schachermayer (1994), we have a sequence $\bar{\nu}^{(n)} \in \text{conv}(\nu^{(n)}, \nu^{(n+1)}, \ldots)$ which converges μ-a.e. to a limit ν^*, since all the $\nu^{(n)}$ are bounded. Now because of the boundedness assumptions, and because $\tilde{g} \geq 0$, it is obvious from (4.4) that the sequence $Y^{(n)}$ is bounded in L^2, and so is uniformly integrable. Using the fact that f was assumed bounded, we deduce the convergence in (4.10). □

Step 2. Write the dual process Y^* in product form as

$$Y_t^* \equiv Z_t^* \beta_t^* \equiv \exp(\int_0^t \psi_s \cdot dW_s - \frac{1}{2}\int_0^t |\psi_s|^2 \, ds). \exp(-\int_0^t r_s^* \, ds)$$

where

$$\begin{aligned} \psi_t &= \sigma_t^{-1}(r_t \mathbf{1} - b_t - \nu_t^*), \\ r_t^* &= r_t + \tilde{g}(t, \nu_t^*). \end{aligned}$$

By the Cameron-Martin-Girsanov theorem (see, for example, Rogers & Williams (2000), IV.38), the martingale Z^* defines a new measure P^* via the recipe $dP^* = Z_T^* \, dP$, and in terms of this

$$dW_t = dW_t^* + \psi_t dt$$

where W^* is a P^*-Brownian motion. The term β^* is interpreted as a stochastic discount factor, and the equality (4.9) can be equivalently expressed as

$$\xi = E^*[\int_0^T \beta_s^* f_s \, ds + \beta_T^* f_T].$$

The bounded P^*-martingale

$$\begin{aligned} M_t &\equiv E^*\Big[\int_0^T \beta_s^* f_s \, ds + \beta_T^* f_T \Big| \mathcal{F}_t\Big] \\ &= \int_0^t \beta_s^* f_s \, ds + E^*\Big[\int_t^T \beta_s^* f_s \, ds + \beta_T^* f_T \Big| \mathcal{F}_t\Big] \end{aligned}$$

has an integral representation

$$M_t = \xi + \int_0^t \theta_s \cdot dW_s^*$$

for some previsible square-integrable integrand θ (see, for example, Rogers & Williams (2000), IV.36). Routine calculations establish that the process

$$X_t^* \equiv (M_t - \int_0^t \beta_s^* f_s \, ds)/\beta_t^*$$

satisfies

$$\begin{aligned}
dX_t^* &= X_t^*[r_t^* dt + \pi_t^* \cdot \sigma_t \, dW_t^*] - f_t dt \\
&= X_t^*[r_t^* dt + \pi_t^* \cdot \sigma_t \, dW_t + \pi_t^* \cdot (b_t - r_t \mathbf{1} + \nu_t^*) dt] - f_t dt \\
&= X_t^*[r_t dt + \pi_t^* \cdot \{\sigma_t \, dW_t + (b_t - r_t \mathbf{1}) dt\} + g(t, \pi_t^*) dt + \varepsilon_t dt] - f_t dt \\
&\equiv X_t^*[dZ_t/Z_t + \varepsilon_t dt] - f_t dt, \qquad (4.11)
\end{aligned}$$

where we have used the notations

$$\begin{aligned}
\pi_t^* &\equiv (\sigma_t^{-1})^T \cdot \theta_t / (\beta_t^* X_t^*), \\
\varepsilon_t &\equiv \tilde{g}(t, \nu_t^*) - g(t, \pi_t^*) + \pi_t^* \cdot \nu_t^* \geq 0, \\
dZ_t &= Z_t[r_t dt + \pi_t^* \cdot \{\sigma_t \, dW_t + (b_t - r_t \mathbf{1}) dt\} + g(t, \pi_t^*) dt].
\end{aligned}$$

We have moreover that $X_0^* = \xi$ and $X_T^* = f_T$, so *provided we could show that ε were zero,* we have constructed a solution pair (c, X) to (4.1) for which $c_s = f_s$ for $0 \leq s < T$, and $X_T = f_T$; *we therefore have the required conclusion $f \in \mathcal{X}(\xi)$ provided $\varepsilon = 0$.*

Step 3. The goal is now clear, and the proof is quite straightforward. If we construct the process X as solution to

$$dX_t = X_t(dZ_t/Z_t) - f_t dt, \qquad X_0 = \xi,$$

then we have from (4.11) that

$$d(X_t^* - X_t) = (X_t^* - X_t) dZ_t/Z_t + \varepsilon_t X_t^* dt$$

and hence

$$X_t^* - X_t = Z_t \int_0^t (\varepsilon_s X_s^*/Z_s) \, ds \geq 0. \qquad (4.12)$$

In particular, since X^* is a bounded process, $-X$ is bounded below by some constant. Now take[8] a process ν such that for all t

$$\tilde{g}(t, \nu_t) = g(t, \pi_t^*) - \pi_t^* \cdot \nu_t,$$

and form the process $Y \in \mathcal{Y}(1)$ from ν according to (4.4); because of the boundedness of r, b, and ν, Y is dominated by an integrable random variable. Repeating the calculation of (4.6) gives us

$$d(X_t Y_t) \doteq -Y_t f_t dt;$$

[8] If $\tilde{g}$ is strictly convex, there is no problem, as the value of $\hat{\nu}$ is unique. More generally we need a measurable selection, but we omit further discussion of this point. In any case, the function g can be uniformly approximated above and below by smooth concave functions, and the result will hold for these; see Section 6 for further discussion.

thus $-X_t Y_t - \int_0^t Y_s f_s \, ds$ is a local martingale bounded below by an integrable random variable, therefore a supermartingale. Hence

$$
\begin{aligned}
\xi &\leq E\Big[\int_0^T Y_s f_s \, ds + Y_T X_T\Big] \\
&\leq E\Big[\int_0^T Y_s f_s \, ds + Y_T X_T^*\Big] && (4.13) \\
&= E\Big[\int_0^T Y_s f_s \, ds + Y_T f_T\Big] \\
&\leq \xi && (4.14)
\end{aligned}
$$

where inequality (4.13) will be strict unless $\int_0^T \varepsilon_s \, ds = 0$ almost surely, from (4.12), and (4.14) is just the definition of ξ. The conclusion that

$$\tilde{g}(t, \nu_t^*) = g(t, \pi_t^*) - \pi_t^* \cdot \nu_t^*$$

μ-a.e. now follows, and so $f \in \mathcal{X}(\xi)$, as required. □

Remarks. We have assumed that $\mathcal{X}(x)$ consists of bounded processes, and have used this boundedness hypothesis in several places. Some such boundedness restriction does appear to be needed in general; however, we expect to argue at the end that no real loss of generality has occurred. For example, in this situation if we were to have taken the larger feasible set $\bar{\mathcal{X}}(x)$ to be the set of *all* optional processes f dominated by some (c, X) solving (4.1), not just the bounded ones, then the new value

$$\bar{u}(x) \equiv \sup_{f \in \bar{\mathcal{X}}(x)} \int U(s, f(s)) \mu(ds)$$

certainly is no smaller than the value $u(x)$ we have been working with. But as we remarked earlier, for any $f \in \bar{\mathcal{X}}(x)$ and $h \in \mathcal{Y}(y)$,

$$\int fh \, d\mu \leq xy,$$

and the inequality analogous to (3.14) holds:

$$\tilde{u}(y) \geq \bar{u}(x) - xy,$$

since the argument leading to (3.11) works just as well for $f \in \bar{\mathcal{X}}(x)$. We therefore have

$$\tilde{u}(y) \geq \bar{u}(x) - xy \geq u(x) - xy;$$

taking the supremum over x, the two ends of these inequalities have been proved to give the same value, so the result holds good for $\bar{u}$ as well.

5 Dual problems made useful

If duality can only turn one impossible problem into another, then it is of no practical value. However, as I shall show in this Section by presenting in full the analysis of the problem of Broadie, Cvitanic & Soner (1998), there really *are* situations where duality can do more than 'reveal the structure' of the original problem, and can in fact lead to a complete solution.

For this problem, introduced in Section 2, it turns out to be more convenient to work with log prices. Since the original share prices satisfy

$$dS_t^i = S_t^i \Big[\sum_j \sigma_{ij} dW_t^j + \rho_i dt \Big],$$

the log prices $X_t^i \equiv \log S_t^i$ satisfy

$$dX_t^i = \sigma_{ij} dW_t^j + b_i dt, \tag{5.1}$$

where $b_i \equiv \rho_i - a_{ii}/2$, $a \equiv \sigma\sigma^T$, and we use the summation convention in (5.1). We further define

$$\psi(X) \equiv \log \varphi(e^X),$$

so that the aim is to super-replicate the random variable $B = \exp(\psi(X_T))$. According to the result of Exercise 6, we must compute

$$\sup_\nu E[Y_T(\nu) B], \tag{5.2}$$

where $Y(\nu)$ solves

$$Y_t^{-1} dY_t = \sigma^{-1}(r\mathbf{1} - \rho - \nu_t) \cdot dW_t - (r + \tilde{g}(\nu_t))dt$$

with initial condition $Y_0 = 1$. But the dual form (5.2) of the problem can be tackled by conventional HJB techniques. Indeed, if we define

$$f(t, X) \equiv \sup_\nu E\Big[\frac{Y_T(\nu)}{Y_t(\nu)} B \;\Big|\; X_t = X \Big],$$

then for any process ν we shall have

$$z_t \equiv Y_t(\nu) f(t, X_t) \quad \text{is a supermartingale, and a martingale for optimal } \nu. \tag{5.3}$$

Abbreviating $Y(\nu)$ to Y, the Itô expansion of z gives us

$$dz_t \doteq Y_t[\ \mathcal{L}f(t, X_t) - (r + \tilde{g}(\nu_t)) f(t, X_t) + \nabla f(t, X_t) \cdot (r\mathbf{1} - \rho - \nu_t)\ ]dt,$$

where $\mathcal{L}$ is the generator of X,

$$\mathcal{L} \equiv \frac{1}{2} a_{ij} D_i D_j + b_i D_i + \frac{\partial}{\partial t},$$

and $D_i \equiv \partial/\partial x_i$. The drift term in z must be non-positive in view of the optimality principle (5.3), so we conclude that

$$\begin{aligned} 0 &= \sup_\nu [\, \mathcal{L}f(t,X_t) - (r+\tilde{g}(\nu))f(t,X_t) + \nabla f(t,X_t)\cdot(r\mathbf{1}-\rho-\nu)\,] \\ &= \mathcal{L}f(t,X_t) + \nabla f(t,X_t)\cdot(r\mathbf{1}-\rho) - rf(t,X_t) \\ &\qquad -f(t,X_t)\inf_\nu\{\tilde{g}(\nu) + \nabla(\log f)(t,X_t)\cdot\nu\} \\ &= \mathcal{L}f(t,X_t) + \nabla f(t,X_t)\cdot(r\mathbf{1}-\rho) - rf(t,X_t) - f(t,X_t)g(\nabla(\log f)(t,X_t)). \end{aligned}$$

For this to be possible, it has to be that

$$\nabla(\log f)(t,x) \in C \quad \text{for all } (t,x), \tag{5.4}$$

and the equation satisfied by f will be

$$0 = \mathcal{L}_0 f(t,X) - rf(t,X),$$

where $\mathcal{L}_0 = \mathcal{L} + (r - a_{ii}/2)D_i$ is the generator of X in the risk-neutral probability. Thus f satisfies the pricing equation, so we have[9]

$$f(t,x) = E^*[\, \exp(\hat{\psi}(X_T))\,|\, X_t = x\,] \tag{5.5}$$

for some function $\hat{\psi}$ such that

$$f(T,x) = \exp(\hat{\psi}(x)).$$

In order that we have super-replicated, we need to have $\hat{\psi} \geq \psi$, and in order that the gradient condition (5.4) holds we must also have

$$\nabla\hat{\psi}(x) \in C \quad \forall x. \tag{5.6}$$

What is the smallest function $\hat{\psi}$ which satisfies these two conditions? If a function ψ_0 satifies the gradient condition (5.6) and is at least as big as ψ everywhere, then for any x and x' the Mean Value Theorem implies that there is some $x'' \in (x,x')$ such that

$$\psi_0(x) - \psi_0(x') = (x-x')\cdot\nabla\psi_0(x'')$$

so

$$\begin{aligned} \psi_0(x) &\geq \psi(x') + \inf_{v\in C}(x-x')\cdot v \\ &= \psi(x') - \delta(x-x'), \end{aligned}$$

where $\delta(v) \equiv \sup\{-x\cdot v : x \in C\}$ is the support function of C. Taking the supremum over x', we learn that

$$\psi_0(x) \geq \Psi(x) \equiv \sup_y\{\psi(x-y) - \delta(y)\}.$$

[9]... using P^* to denote the risk-neutral pricing measure ...

Now clearly $\Psi \geq \psi$ (think what happens when $y = 0$), but we have further that Ψ satisfies the gradient condition (5.6). Indeed, if not, there would be x and x' such that

$$\begin{aligned}\Psi(x') &> \Psi(x) + \sup_{v \in C}(x' - x) \cdot v \\ &= \Psi(x) + \delta(x - x').\end{aligned}$$

However, using the convexity and positive homogeneity of δ, we have

$$\begin{aligned}\Psi(x') &\equiv \sup_y \{\psi(x' - y) - \delta(y)\} \\ &= \sup_z \{\psi(x - z) - \delta(x' - x + z)\} \\ &= \sup_z \{\psi(x - z) - \delta(z) + \delta(z) - \delta(x' - x + z)\} \\ &\leq \Psi(x) + \delta(x - x'),\end{aligned}$$

a contradiction.

This establishes the gradient condition (5.4) for $t = T$, but why should it hold for other t as well? The answer lies in the expression (5.5) for the solution, together with the fact that X is a drifting Brownian motion, because for any h we have

$$\begin{aligned}&\frac{\Big(f(t, x + h) - f(t, x)\Big)}{|h| f(t, x)} \\ &= \frac{E^*[\, |h|^{-1}(\exp(\hat{\psi}(X_T + x + h)) - \exp(\hat{\psi}(X_T + x))) \,|\, X_t = 0\,]}{E^*[\, \exp(\hat{\psi}(X_T + x))) \,|\, X_t = 0\,]} \\ &\to \frac{E^*[\, \nabla\hat{\psi}(X_T + x) \exp(\hat{\psi}(X_T + x))) \,|\, X_t = 0\,]}{E^*[\, \exp(\hat{\psi}(X_T + x))) \,|\, X_t = 0\,]}\end{aligned}$$

as $|h| \to 0$ under suitable conditions. The final expression is clearly in C, since it is the expectation of a random vector which always takes values in the convex set C.

REMARKS. For another example of an explicitly-soluble dual problem, see the paper of Schmock, Shreve & Wystup (2001).

6 Taking stock.

We have seen how the Lagrangian/Hamiltonian/Pontryagin approach to a range of constrained optimisation problems can be carried out very simply, and can be very effective. The recipe in summary is to introduce a Lagrange multiplier process, integrate by parts, and look at the resulting Lagrangian; the story of Section 1 and the examples of Section 2 are so simple that one could present them to a class of MBA students. However, mathematicians ought to grapple with the next two Sections as well, to be convinced that the approach *can* be turned into proof.

What remains? There are many topics barely touched on in these notes, which could readily be expanded to twice the length if due account were to be taken of major contributions so far ignored. Let it suffice to gather here a few remarks under disparate headings, and then we will be done.

1. Links with the work of Bismut. Bismut's (1975) paper and its companions represent a remarkable contribution, whose import seems to have been poorly digested, even after all these years. The original papers were presented in a style which was more scholarly than accessible, but what he did in those early papers *amounts to the same as we have done here.* To amplify that claim, let me take a simple case of his analysis as presented in the 1975 paper, using similar notation, and follow it through according to the recipe of this account. Bismut takes a controlled diffusion process[10]

$$dx = \sigma(t, x, u)dW + f(t, x, u)dt$$

with the objective of maximising

$$E \int_0^T L(t, x, u)\, dt.$$

The coefficients σ, f and L may be suitably stochastic, x is n-dimensional, W is d-dimensional, u is q-dimensional. By the method advanced here, we would now introduce a n-dimensional Lagrange multiplier process[11]

$$dp = bdt + HdW$$

and absorb the dynamics into the Lagrangian by integrating-by-parts; we quickly obtain the Lagrangian[12]

$$\begin{aligned} \Lambda &= E\Big[\int_0^T L(t, x, u)\, dt - [p \cdot x]_0^T + \int_0^T (x \cdot b + \langle H, \sigma(t, x, u)\rangle)\, dt \\ &\qquad + \int_0^T p \cdot f(t, x, u)\, dt\Big] \\ &= E\Big[p_0 \cdot x_0 - p_T \cdot x_T + \int_0^T (x \cdot b + \mathcal{H}(t, x, u))\, dt\Big], \end{aligned}$$

[10]For economy of notation, any superfluous subscript t is omitted from the symbol for a process.

[11]Bismut's notation. H is $n \times d$.

[12]For matrices A and B of the same dimension, $\langle A, B\rangle$ is the L^2-inner product $tr(AB^T)$.

where the Hamiltonian $\mathcal{H}$ of Bismut's account is defined by

$$\mathcal{H}(t,x,u;p,H) \equiv L(t,x,u) + p \cdot f(t,x,u) + \langle H, \sigma(t,x,u) \rangle).$$

Various assumptions will be needed for suprema to be finite and uniquely attained, but our next step would be to maximise the Lagrangian Λ over choice of u, which would lead us to solve the equations

$$\frac{\partial \mathcal{H}}{\partial u} = 0,$$

and maximising Λ over x would lead to the equations

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial x} &= -b, \\ p_T &= 0. \end{aligned}$$

These are the equations which Bismut obtains. The final section of Bismut (1973) explains the relation between the solution obtained by the Lagrangian approach, and the (value-function) solution of the classical dynamic programming approach.

2. Does this dual approach work for all problems? The answer is 'Yes' and 'No', just as it is for the standard dynamic programming approach. We have seen a number of problems where the dual problem can be expressed quite simply, and generally this is not hard to do, but moving from there to an explicit solution can only rarely be achieved[13] (indeed, *just* as in the standard dynamic programming approach, where it is a few lines' work to find the HJB equation to solve, but only rarely can an explicit solution be found.)

During the workshop, Nizar Touzi showed me uncomfortably many examples for which the dual approach described here offered no useful progress; as he stressed, problems where the control affects the volatility of the processes are usually difficult. Here then is an interesting and very concrete example which seems to be hard to deal with.

Example. An investor has wealth process X satisfying

$$dX_t = \theta_t dS_t, \quad X_0 = x,$$

where θ_t is the number of shares held at time t, and S_t is the share price process satisfying

$$dS_t/S_t = \frac{v}{N - \theta_t}\, dW_t, \quad S_0 = 1,$$

where $v > 0$ and $N > 0$. The modelling idea is that there are only N shares in total, and that as the number of shares held by the rest of the market falls, the volatility of the price increases. The agent's objective is to maximise

$$E[\,U(X_T - S_T)\,].$$

[13] But Chow (1997) proposes methods for approximate numerical solution if all else fails.

Introducing Lagrangian semimartingales $d\xi = \xi(adW + bdt)$ for the dynamics of X and $d\eta = \eta(\alpha dW + \beta dt)$ for the dynamics of S, we form the Lagrangian in the usual fashion:

$$
\begin{aligned}
\Lambda &= \sup E\Big[U(X_T - S_T) - ([X\xi]_0^T - \int_0^T Xb\xi \, dt - \int_0^T a\xi \frac{\theta v S}{N-\theta} \, dt) \\
&\qquad + [S\eta]_0^T - \int_0^T S\eta\beta \, dt - \int_0^T \frac{\alpha\eta v S}{N-\theta} \, dt\Big] \\
&= \sup E\Big[\tilde{U}(\xi_T) + X_0\xi_0 - \eta_0 S_0 + \int_0^T \frac{S}{N-\theta} \, (a\xi\theta v - \eta\beta(N-\theta) - \alpha v\eta) \, dt\Big] \qquad (6.1) \\
&= E\Big[\tilde{U}(\xi_T) + X_0\xi_0 - \eta_0 S_0\Big]
\end{aligned}
$$

provided that the dual-feasibility conditions

$$
\begin{aligned}
\xi_T &= \eta_T \\
b &= 0 \\
\alpha\eta &\geq Na \\
v\alpha + N\beta &\geq 0
\end{aligned}
$$

are satisfied. The last two come from inspecting the integral in (6.1); if the bracket was positive for any value of θ in $(0, N)$, then by taking S arbitrarily large we would have an unbounded supremum for the Lagrangian. The form of this dual problem looks quite intractable; the multiplier processes are constrained to be equal at time T, but the bounds on the coefficients of ξ and η look tough. Any ideas?

3. Links to the work of Kramkov-Schachermayer. Anyone familiar with the paper of Kramkov & Schachermayer (1999) (hereafter, KS) will see that many of the ideas and methods of this paper owe much to that. The fact that this paper has not so far mentioned the asymptotic elasticity property which was so important in KS is because what we have been concerned with here is *solely* the equality of the values of the primal and dual problem; the asymptotic elasticity condition of KS was used at the point where they showed that the supremum in the primal problem was attained, and this is not something that we have cared about. The paper KS works in a general semimartingale context, where the duality result (XY) is really very deep; on the other hand, the problem considered in KS is to optimise the expected utility of terminal wealth, so the problem is the simplest one in terms of objective. It is undoubtedly an important goal to generalise the study of optimal investment and consumption problems to the semimartingale setting; it was after all only when stochastic calculus was extended from the Brownian to the general semimartingale framework that we came to understand the rôle of key objects (semimartingales among them!). Such an extension remains largely unfinished at the time of writing.

4. Equilbria. From an economic point of view, the study of the optimal behaviour of a single agent is really only a step on the road towards understanding an equilibrium of many agents interacting through a market. In the simplest situation of a frictionless market without portfolio constraints, the study of the equilibrium is quite advanced; see Chapter 4 of Karatzas & Shreve (1998). However, once we incorporate portfolio constraints of the type considered for example by Cuoco & Liu (2000), it become very difficult to characterise the equilibria of the system. There are already some interesting studies (see Section 4.8 of Karatzas & Shreve (1998) for a list), but it is clear that much remains to be done in this area.

5. Smoothing utilities. In a remark after stating property (U4), I said that the Inada condition at zero was not really needed; the reason is the following little lemma, which shows that we may always uniformly approximate any given utility by one which does satisfy the Inada condition at 0.

Lemma 2 *Let $C \subseteq \mathbb{R}^d$ be a closed convex cone with non-empty interior, and suppose that $U : C \to \mathbb{R} \cup \{-\infty\}$ is concave, finite-valued on $\mathrm{int}(C)$, and increasing in the partial order of C. Assume that the dual cone C^* also has non-empty interior. Then for any $\varepsilon > 0$ we can find $U_\varepsilon, U^\varepsilon : C \to \mathbb{R} \cup \{-\infty\}$ such that*

$$U(x) - \varepsilon \le U_\varepsilon(x) \le U(x) \le U^\varepsilon(x) \le U(x) + \varepsilon$$

for all $x \in \mathrm{int}(C)$, and such that U_ε, U^ε, are strictly concave, strictly increasing, differentiable, and satisfy the Inada condition for any $x \in \mathrm{int}(C)$

$$\lim_{\lambda \downarrow 0} \frac{\partial}{\partial \lambda} U_\varepsilon(\lambda x) = +\infty = \lim_{\lambda \downarrow 0} \frac{\partial}{\partial \lambda} U^\varepsilon(\lambda x).$$

PROOF. Suppose that $\{x_1, \ldots, x_d\} \subseteq C$ is a basis, and that $\{y_1, \ldots, y_d\} \subseteq C^*$ is a basis. Now the functions

$$\begin{aligned} \tilde{u}_+(y) &\equiv \tilde{U}(y) + \frac{\varepsilon}{2} \exp\Big(- \sum_{j=1}^{d} \sqrt{x_j \cdot y} \Big), \\ \tilde{u}_-(y) &\equiv \tilde{U}(y) + \frac{\varepsilon}{2} \exp\Big(- \sum_{j=1}^{d} \sqrt{x_j \cdot y} \Big) - \frac{\varepsilon}{2}, \end{aligned}$$

which sandwich $\tilde{U}$ to within $\varepsilon/2$, are strictly decreasing in y, and are strictly convex. The dual functions, $u_\pm$, are therefore differentiable in $\mathrm{int}(C)$, increasing, and sandwich U to within $\varepsilon/2$. They may fail to be strictly concave, but by considering instead

$$\begin{aligned} u_{++}(x) &\equiv u_+(x) + \frac{\varepsilon}{2}\{1 - \exp\Big(-\sum_{j=1}^{d} \sqrt{x_j \cdot y})\Big)\} \\ u_{--}(x) &\equiv u_-(x) - \frac{\varepsilon}{2}\exp\Big(-\sum_{j=1}^{d} \sqrt{x_j \cdot y}\Big) \end{aligned}$$

we even have the strict concavity as well, and the Inada condition is evident □

Remark. The assumption of non-empty interior for C is not needed; if the interior is empty, we simply drop down to the subspace spanned by C, in which C has non-empty relative interior, and apply the Lemma there. If C contained a linear subspace, then because U is increasing in the order of C, it must be constant in the direction of that subspace, and we can drop down to the quotient space (which now contains no linear subspace) and work there instead.

7 Solutions to exercises.

SOLUTION TO EXERCISE 1. Expressing the expectation of $\int_0^T Y_s dX_s$ in the two ways, we get (assuming that the means of stochastic integrals dW are all zero)

$$E\Big[X_T Y_T - X_0 Y_0 - \int_0^T Y_s\{\alpha_s X_s + \theta_s \sigma_s \beta_s\} ds\Big], \tag{7.1}$$

and

$$E\Big[\int_0^T Y_s\{r_s X_s + \theta_s(\mu_s - r_s \mathbf{1}) - c_s\} ds\Big]. \tag{7.2}$$

The Lagrangian form now is

$$\begin{aligned}
\Lambda(Y) \;\equiv\; & \sup_{X, c\geq 0, \theta} E\Big[\int_0^T U(s, c_s) ds + U(T, X_T) + \int_0^t Y_s\{r_s X_s + \theta_s(\mu_s - r_s \mathbf{1}) - c_s\} ds \\
& \qquad - X_T Y_T + X_0 Y_0 + \int_0^T Y_s\{\alpha_s X_s + \theta_s \sigma_s \beta_s\} ds\Big] \\
=\; & \sup_{X, c\geq 0, \theta} E\Big[\int_0^T \{U(s, c_s) - Y_s c_s\} ds + U(T, X_T) - X_T Y_T + X_0 Y_0 \\
& \qquad + \int_0^T Y_s\{r_s X_s + \theta_s(\mu_s - r_s \mathbf{1}) + \alpha_s X_s + \theta_s \sigma_s \beta_s\} ds\Big]. \qquad (7.3)
\end{aligned}$$

Now the maximisation of (7.3) over $c \geq 0$ and $X_T \geq 0$ is very easy; we obtain

$$\begin{aligned}
\Lambda(Y) \;=\; & \sup_{X\geq 0, \theta} E\Big[\int_0^T \tilde{U}(s, Y_s) ds + \tilde{U}(T, Y_T) + X_0 Y_0 \\
& \qquad + \int_0^T Y_s\{r_s X_s + \theta_s(\mu_s - r_s \mathbf{1}) + \alpha_s X_s + \theta_s \sigma_s \beta_s\} ds\Big],
\end{aligned}$$

where $\tilde{U}(s, y) \equiv \sup_x [U(s, x) - xy]$ is the convex dual of U. The maximisation over $X_s \geq 0$ results in a finite value if and only if the complementary slackness condition

$$r_s + \alpha_s \leq 0 \tag{7.4}$$

holds, and maximisation over θ_s results in a finite value if and only if the complementary slackness condition

$$\sigma_s \beta_s + \mu_s - r_s \mathbf{1} = 0 \tag{7.5}$$

holds. The maximised value is then

$$\Lambda(Y) = E\Big[\int_0^T \tilde{U}(s, Y_s) ds + \tilde{U}(T, Y_T) + X_0 Y_0\Big]. \tag{7.6}$$

The dual problem is therefore the minimisation of (7.6) with the complementary slackness conditions (7.4), (7.5). But in fact, since the dual functions $\tilde{U}(t,\cdot)$ are decreasing, a little thought shows that we want Y to be big, so that the 'discount rate' α will be as large as it can be, that is, the inequality (7.4) will actually hold with equality. This gives the stated form of the dual problem.

Solution to Exercise 2. We use the result of Example 0. Since the dual function $\tilde{u}_0$ of u_0 is just $-u_0$, we have from the dual problem to Example 0, (1.12), that

$$\begin{aligned}\sup E[\, u_0(X_T - B)\,] &= \inf_Y E[\, \tilde{u}_0(Y_T) - Y_T B + xY_0\,] \\ &= \inf_{Y_T \geq 0} E[\, xY_0 - Y_T B\,].\end{aligned}$$

Clearly, this will be $-\infty$ if

$$x < \sup_{Y_T \geq 0} E[\, Y_T B / Y_0\,],$$

and zero else. The statement (2.6) follows.

Solution to Exercise 3a. Introducing the positive Lagrangian semimartingale Y in exponential form

$$dY_t = Y_{t-} dz_t = Y_{t-}(dm_t + dA_t),$$

where m is a local martingale and A is a process of finite variation, and integrating by parts, we find that

$$\begin{aligned}\int_0^T Y_{t-} dX_t &= X_T Y_T - X_0 Y_0 - \int_0^T X_{t-} dY_t - [X,Y]_T \\ &= \int_0^T Y_{t-} H_t dS_t.\end{aligned}$$

Hence the Lagrangian is

$$\begin{aligned}\Lambda(Y) &\equiv \sup E\Big[\, U(X_T) + \int_0^T Y_{t-} H_t dS_t - X_T Y_T + X_0 Y_0 \\ &\qquad\qquad + \int_0^T X_{t-} dY_t + [X,Y]_T\,\Big] \\ &= \sup E\Big[\, U(X_T) - X_T Y_T + X_0 Y_0 + \int_0^T Y_{t-} H_t\, dS_t + \\ &\qquad\qquad \int_0^T X_{t-} Y_{t-}(dm_t + dA_t) + \int_0^T H_t Y_{t-}\, d[m,S]_t\,\Big] \\ &= \sup E\Big[\, \tilde{U}(Y_T) + X_0 Y_0 + \int_0^T Y_{t-}(H_t\,(dS_t + d[m,S]_t) + X_{t-} dA_t)\,\Big]\end{aligned}$$

if means of stochastic integrals with respect to local martingales are all zero. Maximising the Lagrangian over $X \geq 0$, we obtain the dual-feasibility condition that $dA \leq 0$. Next, by maximising over H we see that we must have $dS + d[m, S] \doteq 0$, from which

$$\begin{aligned} d(XY) &= X_- dY + Y_- dX + d[X,Y] \\ &\doteq X_- Y_- dA + Y_- H(dS + d[m,S]) \\ &\doteq X_- Y_- dA \end{aligned}$$

so that XY is a non-negative supermartingale.

Solution to Exercise 3b. Differentiating (2.9) with respect to λ, we find that condition (2.8) is exactly the condition for the derivative to be non-negative throughout $[-1, 1]$. Hence the agent's optimal policy is just to invest all his money in the share. Could $U'(S_1)$ be an equivalent martingale measure? This would require

$$E[U'(S_1)(S_1 - S_0)] = 0,$$

or equivalently,

$$E[\sqrt{S_1}] = E[1/\sqrt{S_1}].$$

It is clear that by altering the $(p_n)_{n \geq 0}$ slightly if necesary, this equality can be broken.

Solution to Exercise 4. Introduce the Lagrangian semimartingale Y satisfying

$$dY_t = Y_t\{\ \alpha_t \cdot \sigma_t dW_t + \beta_t dt\ \} \tag{7.7}$$

and now develop the two different expressions for $\int Y dX$, firstly as

$$\begin{aligned} \int_0^T Y_t dX_t &= Y_T X_T - Y_0 X_0 - \int_0^T X_t\, dY_t - [X,Y]_T \\ &= Y_T X_T - Y_0 X_0 - \int_0^T X_t Y_t \{\alpha_t \cdot \sigma_t dW_t + \beta_t dt\} - [X,Y]_T \\ &= Y_T X_T - Y_0 X_0 - \int_0^T X_t Y_t \{\alpha_t \cdot \sigma_t dW_t + \beta_t dt + \alpha_t \cdot V_t\, \pi_t\, dt\} \\ &\doteq Y_T X_T - Y_0 X_0 - \int_0^T X_t Y_t \{\beta_t + \alpha_t \cdot V_t\, \pi_t\}\, dt. \end{aligned} \tag{7.8}$$

The symbol $\doteq$ signifies that the two sides of the equation differ by a local martingale vanishing at zero. Next, we express $\int Y dX$ as

$$\int_0^T Y_t dX_t = \int_0^T Y_t X_t \Big[\, r_t dt + \pi_t \cdot \{\sigma_t dW_t + (b_t - r_t \mathbf{1}) dt\} + g(t, \pi_t) dt\, \Big]$$

$$
\begin{aligned}
& \qquad\qquad\qquad\qquad - \int_0^T Y_t c_t dt \\
&\doteq \int_0^T Y_t X_t \Big[r_t + \pi_t \cdot (b_t - r_t \mathbf{1}) + g(t, \pi_t) \Big] dt - \int_0^T Y_t c_t dt. \quad (7.9)
\end{aligned}
$$

The Lagrangian form is now

$$
\begin{aligned}
\Lambda \;\equiv\; & E\Big[\int_0^T U(s, c_s) ds + U(T, X_T) \\
& + \int_0^T Y_t X_t \Big[r_t + \pi_t \cdot (b_t - r_t \mathbf{1}) + g(t, \pi_t) \Big] dt - \int_0^T Y_t c_t dt \\
& - Y_T X_T + Y_0 X_0 + \int_0^T X_t Y_t \{ \beta_t + \alpha_t \cdot V_t \pi_t \} \, dt \Big]. \qquad (7.10)
\end{aligned}
$$

Maximising this over X_T and c gives

$$
\begin{aligned}
\Lambda \;=\; & E\Big[\int_0^T \tilde{U}(t, Y_t) \, dt + \tilde{U}(T, Y_T) + Y_0 X_0 \\
& + \int_0^T X_t Y_t \{ \beta_t + \alpha_t \cdot V_t \pi_t + r_t + \pi_t \cdot (b_t - r_t \mathbf{1}) + g(t, \pi_t) \} \, dt.
\end{aligned}
\qquad (7.11)
$$

We find on the way the dual feasibility conditions on Y that, almost surely, $Y_t \geq 0$ for almost every t, and and $Y_T \geq 0$, with strict inequality for t for which U_t is unbounded above.

Since we have that Y and X are both non-negative processes, maximising (7.11) over π amounts to maximising the expression

$$
g(t, \pi) - \pi \cdot (r_t \mathbf{1} - b_t - V_t \alpha_t)
$$

for each t; the maximised value of this expression can be written in terms of the convex dual $\tilde{g}(t, \cdot)$ of g as

$$
\tilde{g}(t, \nu_t),
$$

where ν is related to α by

$$
\nu_t \equiv r_t \mathbf{1} - b_t - V_t \alpha_t. \qquad (7.12)
$$

Alternatively, we can express α in terms of ν as

$$
\alpha_t = V_t^{-1} (r_t \mathbf{1} - b_t - \nu_t). \qquad (7.13)
$$

The value of (7.11) when so maximised over π is therefore

$$
\Lambda = E\Big[\int_0^T \tilde{U}(t, Y_t) \, dt + \tilde{U}(T, Y_T) + Y_0 X_0 + \int_0^T X_t Y_t \{ \beta_t + r_t + \tilde{g}(t, \nu_t) \} \, dt \Big]. \qquad (7.14)
$$

Finally, we consider the maximisation of Λ over X. This leads to the dual-feasiblity condition

$$\beta_t + r_t + \tilde{g}(t, \nu_t) \le 0 \tag{7.15}$$

and a maximised value of the Lagrangian of the simple form

$$\Lambda = E\Big[\int_0^T \tilde{U}(t, Y_t)\, dt + \tilde{U}(T, Y_T) + Y_0 X_0\Big].$$

But $X_0 = x$, so we now believe that the dual problem must be to find

$$\inf_Y E\Big[\int_0^T V(t, Y_t)\, dt + V(T, Y_T) + xY_0\Big]$$

where

$$Y_t^{-1} dY_t = V_t^{-1}(r_t \mathbf{1} - b_t - \nu_t) \cdot \sigma_t dW_t - (r_t + \tilde{g}(t, \nu_t))dt - \varepsilon_t dt,$$

where ε is some non-negative process, from (7.15). Since we are looking to minimise the Lagrangian over Y, and since V_t is decreasing, it is clear that we should take $\varepsilon \equiv 0$, leading to the dynamics

$$Y_t^{-1} dY_t = V_t^{-1}(r_t \mathbf{1} - b_t - \nu_t) \cdot \sigma_t dW_t - (r_t + \tilde{g}(t, \nu_t))dt$$

for Y.

Solution to Exercise 5. According to the machine, the Lagrangian is

$$\begin{aligned}\Lambda \;=\; \sup E\Big[&\int_0^T U(c_t)\, dt + u(X_T, Y_T) + \int_0^T (r_t X_t - c_t)\xi_t\, dt - X_T\xi_T + X_0\xi_0 \\ &+ \int_0^T X_t\xi_t\beta_t\, dt + \int \rho_t Y_t \eta_t\, dt - Y_T\eta_T + Y_0\eta_0 + \int_0^T Y_t\eta_t(b_t + \sigma_t a_t)\, dt \\ &+ \int_0^T ((1-\varepsilon)\xi_t - \eta_t)\, dM_t + \int_0^T (\eta_t - (1+\delta)\xi_t)\, dL_t\Big]\end{aligned}$$

Maximising over increasing M and L, we see that we must have the dual feasibility conditions

$$(1-\varepsilon)\xi_t \le \eta_t \le (1+\delta)\xi_t$$

and the maximised value of the integrals dM and dL will be zero. The maximisation over c and over (X_T, Y_T) is straightforward and transforms the Lagrangian to

$$\begin{aligned}\Lambda \;=\; \sup E\Big[&\int_0^T \tilde{U}(\xi_t)\, dt + \tilde{u}(\xi_T, \eta_T) + X_0\xi_0 + Y_0\eta_0 \\ &+ \int_0^T X_T\xi_T(r_t + \beta_t)\, dt + \int_0^T Y_t\eta_t(\rho_t + b_t + \sigma_t a_t)\Big].\end{aligned}$$

Maximising over X and Y yields the dual feasibility conditions

$$r_t + \beta_t \leq 0 \tag{7.16}$$
$$\rho_t + b_t + \sigma_t a_t \leq 0 \tag{7.17}$$

with the final form of the Lagrangian as

$$E\Big[\int_0^T \tilde{U}(\xi_t)\, dt + \tilde{u}(\xi_T, \eta_T) + X_0\xi_0 + Y_0\eta_0\,\Big].$$

The by now familiar monotonicity argument shows that in trying to minimise this over multipliers (ξ, η) we would have the two dual-feasibility conditions (7.16) and (7.17) satisfied with equality.

Solution to Exercise 6. At least at the heuristic level, which is all we are concerned with for the moment, this exercise follows from Example 4 in the same way that Exercise 2 was derived from Example 0. Just follow through the solution to Exercise 4 assuming that $U(t, .) = 0$ for $t < T$, and $U(T, X_T) = u_0(X_T - \varphi(S_T))$.

References

BISMUT, J.-M. (1973), Conjugate convex functions in optimal stochastic control, *Journal of Mathematical Analysis and Applications* **44**, 384–404.

BISMUT, J.-M. (1975), Growth and optimal intertemporal allocation of risks, *Journal of Economic Theory* **10**, 239–257.

BROADIE, M., CVITANIC, J., and SONER, H. M. (1998), Optimal replication of contingent claims under portfolio constraints, *Review of Financial Studies* **11**, 59–79.

CHOW, G. C. (1997), *Dynamic Economics*, Oxford University Press, New York (ISBN=0195101928).

CUOCO, D. (1997), Optimal consumption and equilibrium prices with portfolio constraints and stochastic income, *Journal of Economic Theory* **72**, 33–73.

CUOCO, D. and CVITANIC, J. (1998), Optimal consumption choices for a "large" investor, *Journal of Economic Dynamics & Control* **22**, 401–436.

CUOCO, D. and LIU, H. (2000), A martingale characterization of consumption choices and hedging costs with margin requirements, *Mathematical Finance* **10**, 355-385.

CVITANIC, J. and KARATZAS, I. (1992), Convex duality in constrained portfolio optimization, *Annals of Applied Probability* **2**, 767–818.

CVITANIC, J. and KARATZAS, I. (1993), Hedging contingent claims with constrained portfolios, *Annals of Applied Probability* **3**, 652–681.

CVITANIC, J. and KARATZAS, I. (1996), Hedging and portfolio optimization under transaction costs: a martingale approach, *Mathematical Finance* **6**, 133–163.

El KAROUI, N. and QUENEZ, M. C., (1991), Programmation dynamique et évaluation des actifs contingents en marché incomplet, *Comptes Rendus de l'Académie des Sciences de Paris, Sér 1,* **313**, 851–854.

El KAROUI, N. and QUENEZ, M. C., (1995), Dynamic programming and pricing of contingent claims in incomplete markets, *SIAM Journal on Control and Optimisation* **33**, 29-66.

El KAROUI, N., PENG, S., and QUENEZ, M. C., (1997), Backwards stochastic differential equations in finance, *Mathematical Finance* **7**, 1–71.

HE, H. and PEARSON, N. D. (1991), Consumption and portfolio policies with incomplete markets and short-sale constraints: the infinite dimensional case, *Journal of Economic Theory* **54**, 259–304.

KARATZAS, I. and KOU, S. (1996), On the pricing of contingent claims under constraints, *Annals of Applied Probability* **6**, 321–369.

KARATZAS, I. and SHREVE, S. E. (1998), *Methods of Mathematical Finance*, Springer, New York (ISBN=0387948392)

KLEIN, I. and ROGERS, L. C. G. (2003), Duality in constrained optimal investment and consumption problems: a synopsis. Preprint.

KORN, R. (1992), Option pricing in a model with a higher interest rate for borrowing than for lending. Preprint.

KRAMKOV, D. and SCHACHERMAYER, W. (1999), The asymptotic elasticity of utility functions and optimal investment in incomplete markets, *Annals of Applied Probability* **9**, 904–950.

JOUINI, E. and KALLAL, H (1995), Arbitrage in securities markets with short-sales constraints *Mathematical Finance* **5**, 197–232.

ROGERS, L. C. G. and WILLIAMS, D. (2000) *Diffusions, Markov Processes and Martingales, Vol. 2*, Cambridge University Press, Cambridge (ISBN=0521775930).

SCHMOCK, U., SHREVE, S. E. & WYSTUP, U. (2002), Valuation of exotic options under shortselling constraints, *Finance & Stochastics* **6**, 143–172.

XU, G.-L. and SHREVE, S. E. (1992) A duality method for optimal consumption and investment under short-selling prohibition. I. General market coefficients, *Annals of Applied Probability* **2**, 87–112.