Duality in constrained optimal investment and consumption problems: a synopsis[1]

I. Klein[2] and L.C.G. Rogers[3]

Abstract

In the style of Rogers (2001), we give a unified method for finding the dual problem in a given model by stating the problem as an unconstrained Lagrangian problem. In a theoretical part we prove our main theorem, Theorem 1, that shows that under a number of conditions the value of the dual and primal problems are equal. The theoretical setting is sufficiently general to be applied to a large number of examples including models with transaction costs, such as Cvitanic & Karatzas (1996) (which could not be covered by the setting in Rogers (2001)). To apply the general result one has to verify the assumptions of Theorem 1 for each concrete example. We show how the method applies for two examples, first Cuoco & Liu (1992) and second Cvitanic & Karatzas (1996).

1 Introduction.

In recent years, there has been a great deal of interest in portfolio optimisation problems of various kinds. The problems are outgrowths of the classical optimal investment/consumption problem dealt with in various forms by Merton (1969), Cox & Huang (1989), Karatzas, Lehoczky & Shreve (1987), and deal with a range of issues where the portfolio may be restricted in some way (Cvitanic & Karatzas (1992), Cuoco & Liu (1998) Xu & Shreve (1992)), or where the objective may be to super-replicate some contingent claim while observing a portfolio constraint, for example, that the holding of the money-market account should never be below some fixed value. See Cvitanic & Karatzas (1993), Cvitanic & Karatzas (1996), Karatzas & Kou (1996) for such papers. We should also mention the work of El Karoui & Quenez (1995) on pricing in incomplete markets.

A common theme of all these papers is to take the original problem, which involves a maximisation over a class of policies, and restate it in terms of the 'dual' problem, which involves a minimisation over some family of measures. Now in

[1]Supported partly by EPSRC grant GR/R03006.

[2]Department of Statistics and Decision Support Systems, University of Vienna, e-mail: Irene.Klein@univie.ac.at

[3]Statistical Laboratory, Wilberforce Road, Cambridge CB3 0WB, GB (phone = +44 1223 766806, fax = +44 1223 337958, e-mail = L.C.G.Rogers@statslab.cam.ac.uk).

most of these examples, it turns out that by stating the problem as a suitable unconstrained Lagrangian problem, the 'dual' problem does indeed turn out to be as given in the paper, though there is no explicit appeal to the Lagrangian method in any of the papers under consideration.

Rogers (2001) presented a unified methodology for deriving the dual problem from the primal, which works in an almost mechanical way on a vast number of examples. The essence of the methodology is to view the stochastic dynamics of the controlled system as constraints, just as in the Pontryagin approach, compare also Bismut (1973), (1975). Introducing a Lagrangian semimartingale, the constraint is absorbed into the Lagrangian form of the problem, converting the left-hand-side by integration by parts, and then an unconstrained maximisation is performed over the variables of the problem (in many typical examples, these would be the consumption rate c_t and the wealth process X_t). Since this maximisation is typically achieved by maximising individually for each t in the time set, it is usually easy to do in closed form, and the resulting dual problem, of minimising over the dual variables, is the dual problem we seek. This approach illuminates the origins of the dual problems, which previously have seemed to arise as the result of some computations starting from the solution.

Having found the dual problem, it remains to prove that the values of the dual and the primal problems are indeed equal, and this is typically where the hard work lies. A general theorem is presented in Rogers (2001) which covers many of the one-dimensional examples from the literature[4].

However, the example Cvitanic & Karatzas (1996)[5] shows 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. In the present paper we close this gap and formulate a setting general enough to cover the transaction costs example of Cvitanic & Karatzas (1996) *as well as* all problems that were already covered by Rogers (2001). The method is sufficiently general not only to cover the two–dimensional example of Cvitanic & Karatzas (including one stock and one bond) but that it could cover also models with transaction costs with d assets, for example Kabanov & Last (2002) and with utility Deelstra, Pham & Touzi (2001). However, one has to verify the conditions of Chapter 3 for each problem and in particular condition (XY) can be quite a challenge. Moreover the method cannot only be applied for diffusion models but for general semimartingale models, compare for instance Rogers (2001), Example 3: Kramkov & Schachermayer (1999).

In the present paper we illustrate the methodology of finding the dual problem

[4] It should be emphasised that the residual difficulty of the problem often lies in the verification of the hypotheses of this theorem.

[5] See also Cvitanic & Wang (2001) where this example was further analyzed.

in full detail for the example Cvitanic & Karatzas (1996). We also present the example Cuoco & Liu (2000) but will refer to Rogers (2001) for details which were already presented there. We will verify for these two examples the conditions that are necessary for our main result, Theorem 1 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. We 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). 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; the argument owes a lot to the argument of Kramkov & Schachermayer (1999) but also needs a certain amount of careful convex analysis.

2 Two examples and their dual formulation

Throughout the paper we use $x \cdot y$ to denote the scalar product of two vectors.

2.1 Example 1: Cuoco-Liu (2000)

The paper of Cuoco & Liu (2000) presents a constrained optimisation problem. The formulation is sufficiently general to include as special cases the problems considered in Cvitanic & Karatzas (1992, 1993), El Karoui, Peng & Quenez (1997) and Cuoco & Cvitanic (1998).

The problem. In an economy with finite time horizon $T > 0$, n risky assets $S^1, \dots, S^n$, and a single riskless asset, the wealth process $(X_t)_{0 \le t \le T}$ of an agent satisfies the dynamics

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

where the various processes have conventional interpretations: r is the riskless rate of return, W is an n-dimensional Brownian motion, b is the n-dimensional rate-of-return process, related to the risky assets by

$$dS_t^i = S_t^i(\sigma_t^i dW_t^i + b_t^i dt), \qquad i = 1, \dots, n,$$

π is the n-dimensional portfolio proportions process, and the adapted process c is the rate of consumption. The symbol $\mathbf{1}$ denotes the n-vector all of whose components are 1, and the initial wealth x_0 is given.

The dynamics (1) are completely conventional, apart from the term involving g, which introduces non-linearity in the following way. We require that the function $g : [0, T] \times \mathbb{R}^n \times \Omega \to (-\infty, \infty)$ satisfies the conditions

(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 $\omega \in \Omega$.

As in Cuoco-Liu we make the following boundedness assumptions:

ASSUMPTION (B): b, r, $\Sigma \equiv \sigma\sigma^T$, Σ^{-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.

We shall habitually omit the third argument from appearances of g. The agent is free to choose a portfolio proportions process π and a consumption process c with the aim of maximising the objective

$$E\Big[\int_0^T U(s, c_s) ds + U(T, X_T) \Big], \tag{2}$$

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[6]. This last assumption means that the utility of negative wealth must be $-\infty$, so only non-negative wealth processes X and consumption processes c are admissible. We shall sometimes write $U_t(c)$ for $U(t, c)$.

THE FIRST STEP: IDENTIFYING THE DUAL PROBLEM. As was explained in the introduction, we now regard the dynamics (1) as a constraint to be satisfied by X, π, and c, and introduce the Lagrangian semimartingale Y satisfying

$$dY_t = Y_t\{ \alpha_t \cdot \sigma_t dW_t + \beta_t dt \} \tag{3}$$

where the previsible processes α and β are to be determined. Then two different expressions for $\int Y\, dX$ have to be developed, one by integration by parts, the second by using the dynamics of X. Then state the Lagrangian and so on (compare also the detailed derivation in Example 2). We refer to Rogers (2001),

[6]The derivative $U'(t, c)$ tends to ∞ as $c \downarrow 0$ and tends to 0 as $c \uparrow \infty$, where U' denotes the derivative with respect to the second variable.

Exercise 4, for the explicit derivation of the dual problem. It is shown there that the dual problem is

$$\inf_Y E\Big[\int_0^T V(t, Y_t)\, dt + V(T, Y_T) + x \cdot Y_0\Big] \tag{4}$$

where the process Y solves

$$Y_t^{-1} dY_t = \Sigma_t^{-1}(r_t \mathbf{1} - b_t - \nu_t) \cdot \sigma_t dW_t - (r_t + \tilde{g}(t, \nu_t))dt \tag{5}$$

for some adapted process ν bounded by γ, and where $\tilde{g}(t, \cdot)$ is the convex dual of $g(t, \cdot)$. This is exactly[7] the dual problem of Cuoco & Liu (2000), Section 4.

2.2 Example 2: Cvitanic & Karatzas (1996)

The problem. This is a simple example incorporating transaction costs. The financial market consists of one riskless asset, the bank account or bond, with price B_t given by

$$dB_t = B_t r_t dt, \qquad B_0 = 1$$

and of one risky asset, the stock, with price per share S_t given by

$$dS_t = S_t \Big[\, \rho_t dt + \sigma_t dW_t \,\Big],$$

for simplicity let $S_0 = 1$. T is the time horizon and $(W_t)_{t \in [0,T]}$ is a standard Brownian motion. The processes σ, ρ, σ^{-1}, r are assumed to be uniformly bounded.

A trading strategy is a pair (L, M) of adapted processes with right continuous, increasing paths and $L(0) = M(0) = 0$, where L_t represents the total amount of funds transferred from bank account to stock, M_t the total amounts transferred from stock to bank account. Given proportional transaction costs $0 < \delta, \varepsilon < 1$ for such transfers and initial holdings x^0, x^1 in bank and stock, the holdings X_t^0 of cash and the holdings X_t^1 of the share at time t satisfy the dynamics

$$\begin{aligned} dX_t^0 &= r_t X_t^0 dt + (1 - \varepsilon) dM_t - (1 + \delta) dL_t - c_t dt \\ dX_t^1 &= X_t^1(\sigma_t dW_t + \rho_t dt) - dM_t + dL_t. \end{aligned} \tag{6}$$

The consumption out of the cash holdings is at rate $c_t \geq 0$. Trading strategies L and M and consumption c are to be chosen such that the processes X_t^0 and X_t^1 satisfy the solvency conditions

$$X_t^0 + (1 - \varepsilon) X_t^1 \geq 0, \qquad X_t^0 + (1 + \delta) X_t^1 \geq 0 \qquad \forall t. \tag{7}$$

[7] We have a term for utility of terminal wealth, which Cuoco & Liu omitted for the sake of simplicity in their formulation.

The set of points $C = \{(x^0, x^1) : x^0 + (1-\varepsilon)x^1 \geq 0, x^0 + (1+\delta)x^1 \geq 0\}$ defines a closed convex cone in $\mathbb{R}^2$.

Suppose the objective is to obtain

$$\sup E\left[\int_0^T U(s, c_s)ds + u(X_T)\right], \tag{8}$$

for some concave utility functions $u : C \to \mathbb{R}$ and $U(t, .) : \mathbb{R}^+ \to \mathbb{R}$ which are assumed to be strictly increasing in the relevant order.

STEP 1: IDENTIFYING THE DUAL PROBLEM. We regard the dynamics (6) as a constraint to be satisfied by X^0, X^1, L, M, c and introduce Lagrangian semimartingales Y^0 and Y^1 satisfying

$$\begin{aligned} dY_t^0 &= Y_t^0(\alpha_t dW_t + \beta_t dt) \\ dY_t^1 &= Y_t^1(a_t dW_t + b_t dt), \end{aligned} \tag{9}$$

where α, β, a and b are to be determined.

We develop two different expressions for $\int Y^0 dX^0 + \int Y^1 dX^1$, firstly by an application of the integration by parts formula, (see, for example, Rogers & Williams (2000), Theorem VI.38.3) and the expression (9) for Y:

$$\begin{aligned} \int_0^T Y_t \cdot dX_t &= Y_T \cdot X_T - Y_0 \cdot X_0 - \int_0^T X_{t-} \cdot dY_t - [X^1, Y^1]_T \\ &= Y_T \cdot X_T - Y_0 \cdot X_0 - \int_0^T X_{t-}^0 Y_t^0 \{\alpha_t dW_t + \beta_t dt\} \\ &\qquad - \int_0^T X_{t-}^1 Y_t^1 \{a_t dW_t + b_t dt\} - \int_0^T X_t^1 Y_t^1 \sigma_t a_t dt \\ &\doteq Y_T \cdot X_T - Y_0 \cdot X_0 - \int_0^T X_t^0 Y_t^0 \beta_t dt - \int_0^T X_t^1 Y_t^1 \{b_t + \sigma_t a_t\} dt \end{aligned} \tag{10}$$

The symbol $\doteq$ signifies that the two sides of the equation differ by a local martingale vanishing at zero. Secondly we use the dynamics of X (6):

$$\begin{aligned} &\int_0^T Y_t^0 dX_t^0 + \int_0^T Y_t^1 dX_t^1 \\ = &\int_0^T Y_t^0 X_t^0 r_t dt + \int_0^T Y_t^0 \{(1-\varepsilon)dM_t - (1+\delta)dL_t - c_t dt\} \\ &+ \int_0^T Y_t^1 X_{t-}^1 \{\sigma_t dW_t + \rho_t dt\} + \int_0^T Y_t^1 \{dL_t - dM_t\} \\ \doteq &\int_0^T Y_t^0 \, \{X_t^0 r_t - c_t\} dt + \int_0^T Y_t^1 X_t^1 \rho_t dt \end{aligned}$$

$$+\int_0^T \{(1-\varepsilon)Y_t^0 - Y_t^1\} dM_t + \int_0^T \{Y_t^1 - (1+\delta)Y_t^0\} dL_t \tag{11}$$

If (X^0, X^1, M, L, c) satisfy the constraints implicit in the dynamics (6), then the two expressions (10) and (11) must agree, so for any such feasible tuple the value of the objective (8) will be the same as the value of the Lagrangian

$$\begin{aligned}
\Lambda \;\equiv\; & \sup E\Big[\int_0^T U(t, c_t)\, dt + u(X_T^0, X_T^1) \\
& + \int_0^T \{r_t X_t^0 - c_t Y_t^0\}\, dt - X_T^0 Y_T^0 + X_0^0 Y_0^0 + \int_0^T X_t^0 Y_t^0 \beta_t\, dt \\
& + \int \rho_t X_t^1 Y_t^1\, dt - X_T^1 Y_T^1 + X_0^1 Y_0^1 + \int_0^T X_t^1 Y_t^1 \{b_t + \sigma_t a_t\}\, dt \\
& + \int_0^T \{(1-\varepsilon)Y_t^0 - Y_t^1\}\ dM_t + \int_0^T \{Y_t^1 - (1+\delta)Y_t^0\}\ dL_t\Big]
\end{aligned} \tag{12}$$

To arrive at this, we have assumed that the means of all stochastic integrals with respect to W will vanish. This needs justification, but recall that the justification will come at the second step; in the first step, we are simply identifying the dual problem which the second step will *prove* is the dual problem.

We now simply maximise (12) over admissible (X^0, X^1, M, L, c), which is very easy. Maximising over increasing M and L, we see that we must have the dual feasibility conditions

$$(1-\varepsilon)Y_t^0 \le Y_t^1 \le (1+\delta)Y_t^0 \tag{13}$$

and the maximised value of the integrals dM and dL will be zero. The maximisation over c and over (X_T^0, X_T^1) is straightforward and transforms the Lagrangian to

$$\begin{aligned}
\Lambda \;=\; & \sup E\Big[\int_0^T V(t, Y_t^0)\, dt + v(Y_T) + X_0 \cdot Y_0 \\
& + \int_0^T X_t^0 Y_t^0 (r_t + \beta_t)\, dt + \int_0^T X_t^1 Y_t^1 (\rho_t + b_t + \sigma_t a_t)\Big],
\end{aligned}$$

where $V(t,z) = V_t(z) = \sup_x\{U(t,x) - xz\}$ and $v(y) = \sup_{(x^0,x^1)}\{u(x^0,x^1) - x\cdot y)\}$

are the convex dual functions of U_t and u, respectively.

Maximising over X^0 and X^1 yields the dual feasibility conditions

$$\beta_t + r_t \;\le\; 0 \tag{14}$$
$$b_t + \sigma_t a_t + \rho_t \;\le\; 0 \tag{15}$$

with the final form of the Lagrangian as

$$E\Big[\int_0^T V(t, Y_t^0)\, dt + v(Y_T) + X_0 \cdot Y_0\Big].$$

A monotonicity argument shows that in trying to minimise this over multipliers (Y^0, Y^1) we would have the two dual-feasibility conditions (14) and (15) satisfied with equality and these processes correspond to the processes Z^0 and Z^1 of Cvitanic & Karatzas(1996) via $Z_t^0 = B_t Y_t^0$ and $Z_t^1 = S_t Y_t^1$.

So we believe that the dual problem must be

$$\inf_Y E\Big[\int_0^T V(t, Y_t^0)\, dt + v(Y_T) + X_0 \cdot Y_0\Big],$$

where Y satisfies the dual feasibility conditions (13) and has the dynamics

$$\begin{aligned} dY_t^0 &= Y_t^0(\alpha_t dW_t - r_t dt) \\ dY_t^1 &= Y_t^1(a_t dW_t - (\rho_t + \sigma_t a_t)dt). \end{aligned}$$

3 The general formulation.

We shall present here a general formulation which applies to a wide range of examples from the literature, in particular to Example 1 and 2 above. The multidimensional version is needed for examples with transaction costs, see for instance Example 2. The formulation is sufficiently general for examples in higher dimension than Example 2 (where the dimension is 2). It could in particular be applied to the transaction costs setting of Kabanov & Last (2002) or rather Deelstra, Pham & Touzi (2001) where the utility comes in. However the verification of the conditions which are given below have to be done for each example and in particular condition (XY) is typically the most difficult.

We give the following general model. Suppose we have some finite measure space $(S, \mathcal{S}, \mu)$ and a closed convex cone C in $\mathbb{R}^d$, with dual cone $C^* \equiv \{y \in \mathbb{R}^d : x \cdot y \geq 0 \ \forall x \in C\}$, both of which we shall assume have non-empty interior. The cone C induces an order on $\mathbb{R}^d$, defined by

$$x \preceq y \ \Leftrightarrow \ y - x \in C. \tag{16}$$

We introduce the notation

$$L_C^0(S, \mathcal{S}, \mu) \equiv \{f : S \to C \mid f \text{ is } \mathcal{S}\text{-measurable}\}$$

usually to be abbreviated to L^0_C. Clearly, L^0_C is a convex cone, closed in the L^0 topology. We shall suppose that for each $x \in C$ we have a convex subset $\mathcal{X}(x)$ of L^0_C, with the properties

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

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

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

(X4) *for some $x_0 \in \mathrm{int}(C)$ the constant function $f : s \mapsto x_0$ is in $\mathcal{X}$,*

where we have used the notation

$$\mathcal{X} \equiv \bigcup_{x \in C} \mathcal{X}(x). \tag{17}$$

in stating (X4). The set $\mathcal{X}$ is not necessarily convex. Notice that because $x_0 \in \mathrm{int}(C)$, for any n there is $\lambda_n > 0$ so large that

$$\lambda_n x_0 - B_n \subseteq C, \tag{18}$$

where $B_n \equiv \{z \in \mathbb{R}^d : \|z\|_\infty \leq n\}$. In other terms,

$$B_n \preceq \lambda_n x_0, \quad \text{equivalently,} \quad B_n \subseteq \lambda_n x_0 - C, \tag{19}$$

so that by (X3), any function taking values in $B_n \cap C$ is in $\mathcal{X}$; in particular, constant functions are in $\mathcal{X}$. Notice also the useful property

$$x \in \mathrm{int}(C) \Rightarrow \exists \delta = \delta(x) > 0 \quad \text{such that} \quad x \cdot y \geq \delta \|y\|_\infty \quad \forall y \in C^*. \tag{20}$$

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

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

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

Once again, the notation

$$\mathcal{Y} \equiv \bigcup_{y \in C^*} \mathcal{Y}(y). \tag{21}$$

will serve in future.

The primal and dual quantities are related by the key polarity property which we state as assumption (XY).

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

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

where we have used the notation

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

An immediate consequence of (XY) is the useful inequality

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

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

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

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

We shall without comment assume that the definition of U has been extended to the whole of $S \times \mathbb{R}^d$ by setting $U(s, x) = -\infty$ if $x \notin C$.

Since we have not assumed that U is differentiable, the gradient of U may not be defined in places. However, the notion of the *supergradient*

$$\partial U(s, x) \equiv \{z : U(s, y) \leq U(s, x) + z \cdot (y - x) \ \ \forall y \in C\}$$

stands in for the gradient (and reduces to it where the function is differentiable.) Because U is finite-valued on $\mathrm{int}(C)$, the supergradient is non-empty there. Because U is $\preceq$-increasing, it follows that $\partial U(s, x) \subseteq C^*$. We require the Inada-type conditions:

(U3) *there exists a measurable map* $\nabla U : S \times \mathrm{int}(C) \to \mathbb{R}^d$ *such that* $\nabla U(s, x) \in \partial U(s, x)$ *for all* s *and* $x \in \mathrm{int}(C)$*, and such that*

$$\nabla U(s, nx_0) \equiv \varepsilon_n(s) \to 0 \quad \mu - \text{a.e.} \tag{22}$$

as $n \to \infty$*, with*

$$\sup_n \int |\varepsilon_n(s)| \, \mu(ds) < \infty. \tag{23}$$

(U4) *there exists some* $x_* \in \mathrm{int}(C)$ *such that the concave function*

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

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

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

Two more mild conditions are imposed:

(U5) *there exists* $\psi \in \mathcal{X}$*, taking values in* $\mathrm{int}(C)$*, such that for all* $\varepsilon \in (0,1)$

$$\sup_{z \in C} |\nabla U(s, z + \varepsilon \psi(s))| \in L^1(S, \mathcal{S}, \mu);$$

(U6) *for each* $s \in \mathcal{S}$*,*

$$|U(s,x)|/|x| \to 0 \quad (|x| \to \infty).$$

One last condition on the utility U is needed, which is most naturally expressed in terms of the convex dual function

$$V(s,y) \equiv \sup_{x \in C} \{U(s,x) - x \cdot y\}, \tag{24}$$

which is evidently convex and $\preceq^*$-decreasing. Regularity of V is assured by the following simple result, whose proof is deferred to the Appendix.

Proposition 1 *For every* $s \in \mathcal{S}$*, for each* $z \in \mathrm{int}(C^*)$ *we have* $V(s,z) < \infty$*, and*

$$V(s,z) = \max_{x \in C} \{U(s,x) - x \cdot z\}.$$

The final condition on U is this:

(U7) *for each* $s \in \mathcal{S}$*,* $-\partial V(s,z)$ *is* $\preceq^*$*-decreasing on* $\mathrm{int}(C^*)$*.*

In more detail, this says that if $z \preceq^* z'$ are two elements of $\mathrm{int}(C^*)$, and $x \in -\partial V(s,z)$, $x' \in -\partial V(s,z')$, then $x' \preceq x$.

In terms of U, we define the functions $u : C \to [-\infty, \infty)$ and $v : C^* \to (-\infty, \infty]$ by

$$u(x) \equiv \sup_{f \in \mathcal{X}(x)} \int U(s, f(s)) \mu(ds) \tag{25}$$

and

$$v(y) \equiv \inf_{g \in \mathcal{Y}(y)} \int V(s, g(s)) \mu(ds). \tag{26}$$

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

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

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

$$\begin{aligned} \int U(s, f(s))\, \mu(ds) &\leq \int \Big[U(s, f(s)) - f(s) \cdot g(s) \Big]\, \mu(ds) + x \cdot y \\ &\leq \int V(s, g(s))\, \mu(ds) + x \cdot y. \end{aligned} \tag{27}$$

Taking $g = g_0$ in this inequality tells us that $u < \infty$, and taking $f = f_0$ tells us that $v > -\infty$.

Theorem 1 *The functions u and v are dual:*

$$v(y) = \sup_{x \in C} [\, u(x) - x \cdot y\,], \tag{28}$$

$$u(x) = \inf_{y \in C^*} [\, v(y) + x \cdot y\,]. \tag{29}$$

Proof. Firstly, notice that part of what we have to prove is very easy: indeed, using the inequality (27), by taking the supremum over $f \in \mathcal{X}(x)$ and the infimum over $g \in \mathcal{Y}(y)$ we have that

$$v(y) \geq u(x) - x \cdot y \tag{30}$$

for any $x \in C$ and $y \in C^*$. 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

$$\Phi(f, g) \equiv \int [\, U(s, f(s)) - f(s) \cdot g(s) \,]\, \mu(ds), \tag{31}$$

and introduce the sets $A_n \subseteq C$ by

$$A_n \equiv C \cap (nx_0 - C).$$

The sets A_n are clearly increasing in n, closed and convex, and are bounded because $\mathrm{int}(C^*)$ is assumed non-empty. In view of (19), for any n we shall have for all large enough m that $B_n \cap C \subseteq A_m$.

We now introduce certain subsets of $L^\infty_C(S, \mathcal{S}, \mu)$ defined in terms of the sets A_n: for each n we define

$$\mathcal{B}_n \equiv \{f \in L^\infty_C(S, \mathcal{S}, \mu) : f(s) \in A_n \;\forall s\}. \tag{32}$$

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 $y \in C^*$, 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 $f \mapsto \int f \cdot g \, d\mu$ 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 and the fact that $\psi \in \mathcal{X}$, and the fact that $U(s, \cdot)$ is $\preceq$-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 $f \in \mathcal{B}_n$

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

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

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

$$\sup_{f \in \mathcal{B}_n} \inf_{g \in \mathcal{Y}(y)} \Phi(f, g) = \inf_{g \in \mathcal{Y}(y)} \sup_{f \in \mathcal{B}_n} \Phi(f, g). \tag{33}$$

From this,

$$\sup_{f \in \mathcal{B}_n} \inf_{g \in \mathcal{Y}(y)} \Phi(f, g) = \inf_{g \in \mathcal{Y}(y)} \int V_n(s, g(s))\, \mu(ds) \equiv v_n(y), \tag{34}$$

say, where

$$V_n(s, z) \equiv \sup\{U(s, x) - z \cdot x : x \in A_n\} \uparrow V(s, z). \tag{35}$$

Consequently, $v_n(y) \leq v(y)$. Now in view of the property (19), we have that

$$\mathcal{B}_n \subseteq \mathcal{X} = \bigcup_{x \in C} \xi^{-1}(x),$$

so

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

The sequence $v_n(y)$ clearly increases with n, so the proof will be complete provided we can prove that

$$\lim_n v_n(y) = v(y), \tag{36}$$

for which we fashion a variant of the argument in Kramkov & Schachermayer (1999); some points require a little care, and we present the proof through a sequence of lemmas whose proofs occupy the Appendix.

Proposition 2 *There exists some sequence (h_n) in $\mathcal{Y}(y)$ such that $h_n \to h$ μ-almost everywhere, and such that*

$$\lim_n \int V_n(s, h_n(s)) \, \mu(ds) = \lim_n v_n(y).$$

The limit h is in $\mathcal{Y}(y)$, by property (Y2).

Proposition 3 *For all $n \leq m$, for all $z \in C^*$, for all $s \in \mathcal{S}$,*

$$V_m(s, \varepsilon_n(s) + z) = V(s, \varepsilon_n(s) + z). \tag{37}$$

Proposition 4 *The family* $\{V(s, \varepsilon_n(s) + g(s))^- : n \in \mathbb{N}, g \in \mathcal{Y}(y)\}$ *is uniformly integrable.*

Let us now complete the proof of Theorem 1, using these results. We have

$$\begin{aligned}
v(y) &\leq \int V(s, h(s))\, \mu(ds) \\
&\leq \liminf_n \int V(s, \varepsilon_n(s) + h(s))\, \mu(ds) \\
&\leq \liminf_n \liminf_{m \geq n} \int V(s, \varepsilon_n(s) + h_m(s))\, \mu(ds) \\
&\leq \liminf_n \liminf_{m \geq n} \int V_m(s, \varepsilon_n(s) + h_m(s))\, \mu(ds) \\
&\leq \liminf_n \liminf_{m \geq n} \int V_m(s, h_m(s))\, \mu(ds) \\
&\leq \limsup_n v_n(y) \\
&\leq v(y)
\end{aligned}$$

using respectively: the definition (26) of v and the fact (Proposition 2) that $h \in \mathcal{Y}(y)$; Fatou's Lemma, Proposition 4, and (22); Fatou's Lemma, Proposition 4; Proposition 3; the fact that V_m is $\preceq$-decreasing; Proposition 2; the fact that $v_n(y) \leq v(y)$ for all n. This chain of inequalities of course establishes (36), which finishes the proof of Theorem 1. $\square$

4 The examples concluded.

Now we complete the analysis of Example 1 and 2 using the theory we have developed in Chapter 3. This closes the gap and proves that the dual problems indeed are as assumed in Chapter 2.

4.1 Example 1: Cuoco-Liu (2000)

The second step: proving duality. In order to use Theorem 1 to prove the dual form of the problem, we have to cast the Cuoco-Liu example in the form of Section 3, and verify the conditions of Theorem 1. The dimension d of the problem is 1, and the convex cone C is $\mathbb{R}^+$.

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[8] 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 (1), for all ω,

$$f(t,\omega) \le c(t,\omega), \qquad (0 \le t < T), \qquad f(T,\omega) \le X(T,\omega). \tag{38}$$

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 is discussed in Rogers (2001), Chapter 4.

Next we define $\mathcal{Y}_0(y)$ to be the set of all solutions to (5) with initial condition $Y_0 = y$. From this we define the set $\mathcal{Y}(y)$ to be the collection of all non-negative adapted processes h such that for some $Y \in \mathcal{Y}_0(y)$

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

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

$$\mathcal{U}((t,\omega),x) = U(t,x),$$

and we shall slightly abuse notation and write U in place of $\mathcal{U}$ 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.

The proof of (X1)-(X4) and (Y1) can be found in Rogers (2001).

(Y2) In view of Assumption (B), we know that

$$\tilde{g}(t,x,\omega) = +\infty \quad \forall |x| > \gamma, \forall t, \omega.$$

Suppose that (h^k) is some sequence in $\mathcal{Y}(y)$ converging in μ-measure; we may (and shall) by passing to a subsequence suppose that the sequence converges μ-almost-everywhere to limit h. The aim is to prove that h is dominated by some element of $\mathcal{Y}_0(y)$.

For each k there exists a process ν^k such that the process Y^k, solution to

$$dY_t^k = Y_t^k \Big[\Sigma_t^{-1}(r_t \mathbf{1} - b_t - \nu_t^k) \cdot \sigma_t dW_t - (r_t + \tilde{g}(t,\nu_t^k))dt \Big], \quad Y_0^k = y, \tag{39}$$

dominates h^k μ-a.e.. We may without loss of generality assume that $|\nu_t^k| \le \gamma$ for all t and ω. To see this, if we define the function

$$\Phi(x) \equiv x \, \frac{\gamma \wedge |x|}{|x|},$$

[8] 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).

and replace the process ν^k by $\bar{\nu}^k \equiv \Phi(\nu^k)$, then Y^k is changed to $\bar{Y}^k$, which agrees with Y^k up to the stopping time

$$\tau^k \equiv \inf\{t : \int_0^t I_{\{|\nu^k_s|>\gamma\}} ds > 0\}.$$

However, Y^k vanishes on $(\tau^k \wedge T, T]$, so we have throughout $[0, T]$ that $Y^k \leq \bar{Y}^k$, and it will be sufficient to prove that some limit of the $\bar{Y}^k$ is in $\mathcal{Y}(y)$ and dominates h.

Since we now have that ν^k_t lies in the compact closed unit ball of radius γ, we can apply Lemma A1.1 of Delbaen & Schachermayer (1994) to deduce that we may find $\theta^k \in \mathrm{conv}(\nu^k, \nu^{k+1}, \ldots)$ which converge μ-a.e. to some limit θ. If we now define $\hat{Y}^k$ to be the solution to (39) with ν^k replaced by θ^k, we have

$$\log(\hat{Y}^k_t/y) = \int_0^t \Sigma_s^{-1}(r_s\mathbf{1} - b_s - \theta^k_s)\cdot dZ_s - \int_0^t (r_s + \tilde{g}(s, \theta^k_s) + \frac{1}{2}\theta^k_s \cdot \Sigma_s^{-1}\theta^k_s)\, ds. \quad (40)$$

Passing to a subsequence if necessary, and using the uniform boundedness of the θ^k, we may suppose that the stochastic integral terms on the right of (40) converge uniformly almost surely. We may not be able to deduce the convergence of the Lebesgue integral terms involving $\tilde{g}$, but the Fatou inequality works the right way for us, and yields

$$\begin{aligned}\limsup \log(\hat{Y}^k_t/y) \quad &\leq \quad \int_0^t \Sigma_s^{-1}(r_s\mathbf{1} - b_s - \theta_s)\cdot dZ_s \\ &\qquad - \int_0^t (r_s + \tilde{g}(s, \theta_s) + \frac{1}{2}\theta_s \cdot \Sigma_s^{-1}\theta_s)\, ds \\ &\equiv \quad \log(\hat{Y}_t/y), \end{aligned} \quad (41)$$

say. Now $\hat{Y} \in \mathcal{Y}_0(y)$, and if $\theta^k = \sum_{j\geq k} p^k_j \nu^j$ represents θ^k as a finite convex combination of the ν^j, by the argument that establishes (Y1) we have

$$\sum_{j\geq k} p^k_j h^j_t \leq \sum_{j\geq k} p^k_j Y^j_t \leq \hat{Y}^k_t. \quad (42)$$

But the limit of the leftmost expression in (42) is μ-a.e. equal to h, and the limit superior of the rightmost expression in (42) is at most $\hat{Y}_t$, which establishes that $h \in \mathcal{Y}(y)$, and hence gives property (Y2).

The properties of the utility, and the finiteness assumption are dealt with as follows: properties (U1) and (U2) are evident, moreover $x \mapsto U(s, x)$ is differentiable and (U7) holds. 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)–(U6) are satisfied. For the finiteness condition (F), we must once again check this for each particular case.

The proof of (XY) is the tricky part and is done in Rogers (2001), Chapter 4. This (and the corresponding part of Example 2) requires a condensation argument to construct a limiting optimal process; such arguments are presented in much more general settings in the paper of Guasoni (2002), and various other papers referred to there. However, the presence of running consumption in the objective is a point of difference between our examples and the situations studied in those references, which means that their results cannot be immediately carried over to the present context.

REMARK: CONVEX CONSTRAINTS. If we take the Cuoco-Liu problem and impose the constraint that the portfolio should always lie in some closed convex set K, then the dynamics are still represented by (1), where now g is modified to be $-\infty$ off the set K. However, such a modified g no longer satifies the global Lipschitz condition needed for the proof of the duality relationship. While it is easy to think of ways of approximating the modified g by globally Lipschitz g to which our result *does* apply, establishing that the duality relation holds in the limit appears to be impossible at the level of generality that we have worked at so far. For example, Cuoco & Liu (1998) require global growth conditions on the derivative of U, and that the original g is globally Lipschitz. The details of passing from the original g to the portfolio-constrained g must depend on context, just as establishing (XY) must depend on context, and it seems pointless to try to frame a set of sufficient conditions.

4.2 Example 2: Cvitanic, Karatzas (1996)

STEP 2: PROVING DUALITY.

The finite measure space $(S, \mathcal{S}, \mu)$ is

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

Define a cone $C \subseteq \mathbb{R}^2$ as follows, x denotes the vector $\begin{pmatrix} x^0 \\ x^1 \end{pmatrix}$,

$$C = \left\{ x : \quad x^0 + (1-\varepsilon)x^1 \geq 0, \quad x^0 + (1+\delta)x^1 \geq 0 \right\}. \tag{43}$$

For each $x \in C$ the set $\mathcal{X}^0(x)$ is the collection of all optional $X : S \to C$ fulfilling dynamics (6) with starting value $X_0 = x$. Define $\mathcal{X}(x)$ to be the set of all optional $f = (f^0, f^1)^t : S \mapsto C$ such that for some $X \in \mathcal{X}^0(x)$, for all ω,

$$f(t,\omega) \preceq (c(t,\omega),0)^t \quad (0 \le t < T), \quad f(T,\omega) \preceq X(T,\omega), \tag{44}$$

where $\preceq$ is the partial order induced by the cone C.

The dual cone C^* is given by the following, where $y = \begin{pmatrix} y^0 \\ y^1 \end{pmatrix}$,

$$C^* = \left\{ y : \quad (1+\delta)y^0 \ge y^1 \ge (1-\varepsilon)y^0 \right\}. \tag{45}$$

For $y \in C^*$, we define $\mathcal{Y}_0(y)$ to be the set[9] of all processes $Y = (Y^0, Y^1)^t$ which can be represented for some bounded previsible processes a and α as solutions to the SDEs

$$\begin{aligned} dY_t^0 &= Y_t^0(\alpha_t dW_t - r_t dt) \\ dY_t^1 &= Y_t^1(a_t dW_t - (\rho_t + \sigma_t a_t)dt) \end{aligned} \tag{46}$$

with initial conditions $Y_0^0 = y^0$, $Y_0^1 = y^1$, and such that $Y_t \in C^*$ for all t.

Let $\mathcal{Y}_1(y)$ be all optional $g : S \to C^*$ such that there is $Y \in \mathcal{Y}_0(y)$ with $g \preceq^* Y$, where $\preceq^*$ is the partial order induced by the cone C^*. We then take $\mathcal{Y}(y)$ to be the closure of $\mathcal{Y}_1(y)$ with respect to convergence in μ.

Remark. We take here the closure in $L^0(\mu)$ of the convex set $\mathcal{Y}_1(y)$ since it is not clear whether the condition (Y2) is satisfied for such a set of processes dominated by exponential semimartingales in $\mathcal{Y}_0(y)$. However, $\mathcal{Y}(y)$ remains convex, satisfies (Y2) by definition, and by Fatou's lemma

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

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.

(X1), (X2) and (X3) are trivial.

(X4) By taking $(M, L) \equiv (0,0)$ and $c \equiv 0$ and using the fact that r is bounded, we see from the dynamics of X that, for some small enough $\gamma > 0$, the constant function $(t,\omega) \mapsto (\gamma, 0)$ is in $\mathcal{X} = \cup_{x \in C} \mathcal{X}(x)$ and it is clear that $(\gamma, 0) \in \operatorname{int}(C)$.

(Y1) The proof of the convexity of $\mathcal{Y}(y)$ is similar to the proof of property (X1) in the Cuoco-Liu example, compare Rogers (2001), Chapter 4.

[9]... which is non-empty, as may be seen by taking $\alpha = a = (r-\rho)/\sigma$.

Remark. An analogous proof shows that for $y \in C^*$ and $z \in C^*$ we have that

$$\mathcal{Y}(y) + \mathcal{Y}(z) \subseteq \mathcal{Y}(y+z) \tag{47}$$

(Y2) is trivial as we have defined $\mathcal{Y}(y)$ as a closure.

Before we prove condition (XY) which is the difficult part of this section we say a few words about the utility conditions. Assume we are given the utilities $U(t,.) : \mathbb{R}^+ \to \mathbb{R}$ and $u : C \to \mathbb{R}$ as above (8); as in Cvitanic & Karatzas (1996), we suppose that for each $t \in [0,T]$ the map $x \mapsto U(t,x)$ is a strictly increasing, strictly concave, continuously differentiable function that satisfies the Inada conditions. We shall make the convention that $U(t,x) = -\infty$ for $x < 0$.

We now aim at defining a utility $\mathcal{U} : S \times C \to \mathbb{R} \cup \{-\infty\}$ that fits the general theory (Section 3) and, moreover, gives exactly the objective we want. That is, for f fulfilling (44), we need

$$\int \mathcal{U}(t,f)d\mu \leq E\left[u(X_T^0, X_T^1) + \int_0^T U(s,c_s)ds\right].$$

Construction of $\mathcal{U}$: the linear map $L : C \to \mathbb{R}^2_+$ defined by the matrix

$$\begin{pmatrix} 1 & 1+\delta \\ 1 & 1-\varepsilon \end{pmatrix}$$

is a linear bijection of C with $\mathbb{R}^2_+$ that respects the cone orderings. We propose to define a map $\Phi : \mathbb{R}^2_+ \to \mathbb{R}_+$ by

$$\Phi(x,y) = \sqrt{x^2+y^2}\varphi(\theta)$$

(where $\theta = \tan^{-1}(y/x)$) in such a way that the following properties are satisfied:

(u1) Φ is strictly increasing on $\mathbb{R}^2_+$;

(u2) Φ is concave;

(u3) $\nabla U(\Phi)$ is onto $(0,\infty)^2$.

(u4) $U(\Phi)$ is concave.

It is readily verified that by choosing the function

$$\varphi(x) = \frac{1+\sqrt{4\sin x \cos x}}{2+\sqrt{2}}$$

all the properties (u1)-(u4) hold. Hence the map $\mathcal{U} : [0,T) \times C \to \mathbb{R}$ defined by

$$\mathcal{U}(t,x) = U(t, \Phi(Lx)) \tag{48}$$

is concave, strictly $\preceq$-increasing on C, and is easily seen to have the property that

$$\mathcal{U}(t,(c,0)) = U(t,c). \tag{49}$$

We complete the definition of $\mathcal{U}$ by specifying that

$$\mathcal{U}(T,x) = u(x). \tag{50}$$

Then, for $f \in \mathcal{X}(x)$ we have that

$$\int \mathcal{U}(s,f)d\mu = E\left[\mathcal{U}(T,f_T) + \int_0^T \mathcal{U}(s,f_s)ds\right] \leq E\left[u(X_T) + \int_0^T U(s,c_s)ds\right],$$

as (44) is fulfilled and (49) and (50) hold. And this is exactly what we need, compare with the objective (8).

Now the conditions (U1) to (U7) have to be checked, as well as the finiteness condition (F). Some uniformity in t needs to be assumed for U; sufficient is the condition that for each $a > 0$,

$$\int_0^T U'(s,a)ds < \infty. \tag{51}$$

(U1), (U2) are trivial.

(U3) Taking $x_0 = (1,0)$, it is a simple application of the chain rule that

$$\nabla\mathcal{U}(t,x) = U'(t,\Phi(Lx))\nabla\Phi(Lx)L. \tag{52}$$

Since $\nabla\Phi(nx_0)$ is uniformly bounded, (U3) follows easily by monotone convergence from (51).

(U4) If condition (U4) holds for U, then $\mathcal{U}$ will inherit it.

(U5) This follows quickly from (51) when we take $\psi = (1,0)$.

(U6) will follow if for each $t \in [0,T]$

$$\lim_{x\uparrow\infty} |U(t,x)|/|x| = 0,$$

which we assume is satisfied.

(U7) For this, we have to show that if $z \preceq^* z' \in \mathrm{int}(C^*)$ and $x, x' \in C$ are such that $z \in \partial U(s,x)$, $z' \in \partial U(s,x')$, then $x' \preceq x$. Since any $z \in \mathrm{int}(C^*)$ can be represented as $z = (a,b)L$ for some positive a, b, from the chain rule expression (52) we see that it will be sufficient to show that the relation

$$U'(t, \Phi(x))\nabla\Phi(x) = (a,b) \tag{53}$$

between $x \in \mathbb{R}^2_+$ and $(a,b) \in (0,\infty)^2$ is decreasing. Now it is clear that given $(a,b) \in (0,\infty)^2$ there exists a unique $x \in \mathbb{R}^2_+$ such that (53) holds, because the ratio a/b determines the angular part of x, and then the distance of x from the origin is derived easily, using the Inada condition and the fact that U is C^1. To see that the relation is decreasing, we have to show that if either component of x increases, then *neither* component of (a,b) increases. In terms of derivatives, this condition is

$$\begin{aligned} U''(\Phi)\Phi_x^2 + U'(\Phi)\Phi_{xx} &\leq 0, \\ U''(\Phi)\Phi_x\Phi_y + U'(\Phi)\Phi_{yx} &\leq 0, \\ U''(\Phi)\Phi_y^2 + U'(\Phi)\Phi_{yy} &\leq 0. \end{aligned} \tag{54}$$

But the concavity of Φ and U together with their monotonicity deals with the first and the last, and the middle condition is satisfied because

$$\Phi_{yx}^2 \leq \Phi_{xx}\Phi_{yy}$$

by concavity of Φ, and the consequent inequalities

$$|\Phi_{yx}U'(\Phi)|^2 \leq \Phi_{xx}\Phi_{yy}U'(\Phi)^2 \leq (-U''(\Phi)\Phi_x^2)(-U''(\Phi)\Phi_y^2),$$

the last by the first and last parts of (54).

(XY)

In order to get condition (XY) we will need an analogue of Theorem 3.2 of Kabanov & Last (2002) for our setting, see Proposition 5 below. The proof consists in reproving some results of Kabanov & Last (2002), this is rather straightforward but not completely trivial. The proofs are assembled in a technical report of the ISDS: Some details of Example 2 in Klein & Rogers (2005), Klein (2005). Recall the definition of $\Psi(f) = \{x \in C : f \in \mathcal{X}(x)\}$.

Proposition 5

$$\begin{aligned} \Psi(f) &= \cap_{y \in C^*}\{x \in C : \int f \cdot g \, d\mu \leq x \cdot y \quad \forall g \in \mathcal{Y}(y)\} \\ &= \cap_{y \in C^*}\{x \in C : \sup_{g \in \mathcal{Y}(y)} \int f \cdot g \, d\mu \leq x \cdot y\} \end{aligned}$$

Proof. The proof is in principle reproving Theorem 3.2 of Kabanov & Last (2002) for this setting. The details can be found in Klein (2005).

Proposition 6 *Let $f \in \mathcal{X}$. Then condition (XY) holds, that is*

$$\sup_{g\in\mathcal{Y}(y)} \int f \cdot g \, d\mu = \inf_{x\in\Psi(f)} x \cdot y \tag{55}$$

Proof.

As $X \cdot Y + \int cY^0 dt \geq 0$ for all $X \in \mathcal{X}_0(x)$ and $Y \in \mathcal{Y}_0(x)$ it is rather straightforward to show that $X \cdot Y + \int cY^0 dt$ is a supermartingale. Hence

$$E[X_T \cdot Y_T + \int_0^T c_t Y_t dt] \leq x \cdot y$$

Moreover, for elements $f \in \mathcal{X}(x)$ we have that (44) holds and each $g \in \mathcal{Y}(y)$ is a limit of elements of $\mathcal{Y}_1(y)$ (which are dominated by elements of $\mathcal{Y}_0(y)$). So we learn that the left-hand side of (55) is no greater than the right-hand side, by the definition of the measure μ and by Fatou.

For the proof of the reverse inequality, denote $\Psi_y(f) = \{x \in C : \sup_{g\in\mathcal{Y}(y)} \int f \cdot g d\mu \leq (x, y)\}$, hence, by Proposition 5,

$$\Psi(f) = \cap_{y\in C^*} \Psi_y(f).$$

Observe that $\lambda\mathcal{Y}(y) = \mathcal{Y}(\lambda y)$ for $y \in C^*$ and $\lambda > 0$, so that $\Psi_y(f) = \Psi_{\lambda y}(f)$. This means, it is enough to consider $y = (y^0, y^1)^t \in \tilde{C^*} = \{y \in C^* : y^1 = 1\}$, therefore

$$\Psi(f) = \cap_{y\in\tilde{C^*}} \Psi_y(f). \tag{56}$$

Defining $A_y = \sup_{g\in\mathcal{Y}(y)} \int f \cdot g \, d\mu$, we have

$$\Psi_y(f) = \{x \in C : x \cdot y \geq A_y\}$$

so by Proposition 5, $\Psi(f)$ is an intersection of closed half-spaces.

Assume now that (XY) does not hold, that is, there is $y \in \tilde{C^*}$ such that

$$A_y < \inf_{x\in\Psi(f)} x \cdot y.$$

$\Psi(f)$ is a closed, convex set in $\mathbb{R}^2$. There is $x_y \in \Psi(f)$ with $x_y \cdot y = \inf_{x\in\Psi(f)} x \cdot y$ and

$$A_y < x_y \cdot y.$$

It is clear that x_y has to lie in the boundary $\partial\Psi(f)$ of the set $\Psi(f)$. Now it's a geometric argument. We distinguish three cases:

A) There is a unique tangent to $\partial\Psi(f)$ in x_y and it touches $\partial\Psi(f)$ in more than one point. That's the trivial case, by (56) there is only one possibility for $\Psi(f)$ to look like that, and that is $x_y \cdot y = A_y$, a contradiction.

B) There is a unique tangent to $\partial\Psi(f)$ in x_y and it touches $\partial\Psi(f)$ only in x_y. By (56) there exist sequences y_n and $z_n \in \tilde{C^*}$ with $\lim y_n = \lim z_n = y$ and $y_n^0 < y^0$ and $z_n^0 > y^0$ such that $\lim A_{y_n} = x_y \cdot y$ and $\lim A_{z_n} = x_y \cdot y$. For each n we can write y as convex combination $y = \lambda_n y_n + (1-\lambda_n)z_n$. Then, for a subsequence such that λ_n converges, still $\lim(\lambda_n A_{y_n} + (1-\lambda_n)A_{z_n}) = x_y \cdot y$. On the other hand

$$
\begin{aligned}
& \lambda_n A_{y_n} + (1-\lambda_n)A_{z_n} \\
= \; & \sup_{g\in\mathcal{Y}(\lambda_n y_n)} \int f \cdot g \, d\mu + \sup_{h\in\mathcal{Y}((1-\lambda_n)z_n)} \int f \cdot h \, d\mu \\
= \; & \sup_{k\in[\mathcal{Y}(\lambda_n y_n)+\mathcal{Y}((1-\lambda_n)z_n)]} \int f \cdot k \, d\mu \\
\leq \; & \sup_{g\in\mathcal{Y}(y)} \int f \cdot g \, d\mu = A_y \\
< \; & x_y \cdot y,
\end{aligned}
$$

where the first equation comes from the fact that $\lambda\mathcal{Y}(y) = \mathcal{Y}(\lambda y)$ and the inequality in line 4 comes from (47), i.e., $\mathcal{Y}(y) + \mathcal{Y}(z) \subseteq \mathcal{Y}(y+z)$. But that's a contradiction, for $n \to \infty$.

C) The tangent to $\partial\Psi(f)$ in x_y is not unique. As $\partial\Psi(f)$ is convex there is a tangent from the left hand side and a tangent from the right hand side. By (56) there are y_1 and y_2 in $\tilde{C^*}$ defining the two tangents and such that $y_1^0 < y^0 < y_2^0$ (i.e. such that the tangents cross the y-axes in $x_y \cdot y_1$, $x_y \cdot y_2$ respectively). Moreover, there exist sequences y_n and $z_n \in \tilde{C^*}$ with $\lim y_n = y_1$ and $\lim_n z_n = y_2$ such that $\lim A_{y_n} = x_y \cdot y_1$ and $\lim A_{z_n} = x_y \cdot y_2$. We can write $y = \lambda y_1 + (1-\lambda)y_2$. Choose subsequences such that $y = \lambda_n y_n + (1-\lambda_n)z_n$ and $\lim \lambda_n = \lambda$. Then

$$\lim(\lambda_n A_{y_n} + (1-\lambda_n)A_{z_n}) = \lambda(x_y \cdot y_1) + (1-\lambda)(x_y \cdot y_2) = x_y \cdot y.$$

As before this leads to a contradiction. And so we found that $A_y = x_y \cdot y$ for all $y \in C^*$, which gives (XY). $\square$

5 Appendix: Proofs of auxiliary results.

Proof of Proposition 1.The point to prove, of course, is that there cannot be a sequence (x_n) of points tending to infinity in C at which ever larger values of $U(s,x) - (x,z)$ are attained. So suppose that for some s this were not the case, and that there exist $x_n \in C$, $|x_n| \to \infty$, such that

$$a \leq U(x_n) - x_n \cdot z \tag{57}$$

where for notational simplicity we temporarily drop the explicit dependence on s, and a is some finite real less than $V(z)$. Passing to a subsequence if necessary, we may suppose that $x_n/|x_n| \to \xi \in C \cap S^{d-1}$; since $z \in \mathrm{int}(C^*)$, it must be that $(\xi, z) = 2\varepsilon > 0$. Now we use property (U6); for all large enough n, $|U(x_n)| \leq \varepsilon |x_n|$. But this is incompatible with the inequality (57). □

Proof of Proposition 2. From (34) we may take $g_n \in \mathcal{Y}(y)$ such that

$$\int V_n(s, g_n(s))\, \mu(ds) \leq v_n(y) + n^{-1},$$

and hence by Lemma A1.1 of Delbaen & Schachermayer (1994) there can be found convex combinations $h_n \in \mathrm{conv}(g_n, g_{n+1}, \ldots)$ which converge μ-almost everywhere.

In more detail, since the interior of C is non-empty, there exist $x_1, x_2, \ldots, x_d \in \mathrm{int}(C)$ which are linearly independent. Since for each $j = 1, 2, \ldots, d$ the sequence $x_j \cdot g_n$ consists of non-negative functions, by applying Lemma A1.1 of Delbaen & Schachermayer (1994), we can find a sequence $g'_n \in \mathrm{conv}(g_n, g_{n+1}, \ldots)$ such that $x_1 \cdot g'_n$ converges μ-almost everywhere to a limit in $[0, \infty]$. Applying the same result again, we can make a sequence $g''_n \in \mathrm{conv}(g'_n, g'_{n+1}, \ldots)$ such that $x_j \cdot g'_n$ converges μ-almost everywhere for $j = 1, 2$. Proceeding in the same way, we end up with the required sequence (h_n), such that $x_j \cdot h_n$ is μ-almost everywhere convergent to a limit in $[0, \infty]$ for each $j = 1, 2, \ldots, d$. In view of property (19), Fatou's Lemma, and (XY), the limits must be μ-almost surely finite. This implies that there exists $h_n \in \mathcal{Y}(y)$ converging μ-a.e. to h, which is in $\mathcal{Y}(y)$ using property (Y2), and for which

$$\int V_n(s, h_n(s))\, \mu(ds) \leq \sup_{m \geq n} [\, v_m(y) + m^{-1}\,].$$

□

Proof of Proposition 3.For notational simplicity, we omit the appearance of s. It is a simple result on dual functions that $z \in \partial U(x)$ if and only if $-x \in \partial V(z)$, and then

$$V(z) = U(x) - x \cdot z. \tag{58}$$

Hence $-nx_0 \in \partial V(\varepsilon_n)$. Moreover, $\varepsilon_n \in \mathrm{int}(C^*)$, for if not, there would be some $\tilde{x} \in C$ such that $\tilde{x} \cdot \varepsilon_n = 0$, and then the inequality

$$U(nx_0 + t\tilde{x}) \leq U(nx_0) + t(\tilde{x} \cdot \varepsilon_n) = U(nx_0)$$

contradicts the strict $\preceq$-increase of U.

Using property (U7), for any $x' \in -\partial V(\varepsilon_n + z)$ it must be that $x' \preceq nx_0$, and so $x' \in A_n$. The proposition will be proved if we can show that

$$V(\varepsilon_n + z) \equiv \sup_{x \in C}\{\, U(x) - x \cdot (\varepsilon_n + z)\,\} = U(x') - x' \cdot (\varepsilon_n + z),$$

which is simply an application of (58). □

Proof of Proposition 4.This proof is a slight modification of Lemma 3.2 of Kramkov & Schachermayer (1999), exploiting condition (U4). Firstly we note that

$$\begin{aligned} -V(s,z) &\equiv \inf_{x \in C}\{\, x \cdot z - U(s,x)\,\} \\ &\leq \inf_{\lambda > 0}\{\, (\lambda x_*) \cdot z - U(s, \lambda x_*)\,\} \\ &\leq \inf_{\lambda > 0}\{\, \lambda(x_* \cdot z) - \underline{u}(\lambda)\,\} \\ &\equiv \psi(x_* \cdot z), \end{aligned}$$

where the concave increasing function ψ is the dual function of $\underline{u}$. 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

$$\begin{aligned} \int \varphi(V(s, \varepsilon_n(s) + g(s))^-)\mu(ds) &\leq \int \varphi(\max\{0, \psi(x_* \cdot (\varepsilon_n(s) + g(s)))\}\mu(ds) \\ &\leq \varphi(0)\mu(S) + \int \varphi(\psi(x_* \cdot (\varepsilon_n(s) + g(s))))\mu(ds) \\ &= \varphi(0)\mu(S) + \int x_* \cdot (\varepsilon_n(s) + g(s))\mu(ds). \end{aligned}$$

This is bounded by a finite constant independent of n and g, in view of (23), the fact that the constant function x_* is in $\mathcal{X}$, and property (XY). □

There is a useful little corollary of this proposition.

Corollary 1 *For each $y \in C^*$, there is some $g \in \mathcal{Y}(y)$ for which the infimum defining $v(y)$ in (26) is attained.*

Differentiability of U implies strict convexity of V, which in turn implies uniqueness of the minimising g.

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

$$v(y) \leq \int V(s, g_n(s))\mu(ds) \leq v(y) + n^{-1}. \tag{59}$$

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 (59) Now by Proposition 4 and Fatou's lemma,

$$v(y) \leq \int V(s, g(s))\mu(ds) \leq \liminf_n \int V(s, g_n(s))\mu(ds) \leq v(y),$$

as required. The uniqueness assertion is immediate. $\square$

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.

COX, J.C. and HUANG, C.F. (1989), Optimal consumption and portfolio policies when asset prices follow a diffusion process *Journal of Economic Theory* **49**, 33–83.

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.

CVITANIC, J. and WANG, H. (2001), On optimal terminal wealth under transaction costs, *Journal of Mathematical Economics* **35**, 223–232.

DEELSTRA, G., PHAM, H. and TOUZI, N. (2001), Dual formulation of the utility maximization problem under transaction costs, *Annals of Applied Probability* **11**, 1353–1383.

DELBAEN, F. and SCHACHERMAYER, W. (1994), A general version of the fundamental theorem of asset pricing, *Mathematische Annalen* **123**, 463–520.

El KAROUI, N. and QUENEZ, M. C., (1991), Programmation dynamique et évaluation des actifs contingents en marché incomplet, *Comptes Rendus de l'Académie des Sci-*

ences 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.

GUASONI, P., (2002), Optimal investment with transaction costs and without semimartingales, *Ann. Appl. Prob.* **4**, 1227–1246

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.

KABANOV, Y.M. and LAST, G. (2002), Hedging under transaction costs in currency markets: a continuous–time model, *Mathematical Finance* **12**, 63–70.

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

KARATZAS, I., LEHOCZKY, J.P. and SHREVE, S.E. (1987), Optimal portfolio and consumption decisions for a "small investor" on a finite horizon, *SIAM Journal of Control and Optimisation* **25**, 1557–1586.

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

KLEIN, I. (2005), Some Details of Example 2 in Klein & Rogers (2005), *Technical Report of the ISDS http://homepage.univie.ac.at/irene.klein/cktech05.pdf*

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.

MERTON, R.C., (1969), Lifetime portfolio selection under uncertainty: the continuous-time case, *Rev. Econom. Statist.* **51**, 247–357

ROGERS, L. C. G. (2001), Duality in constrained optimal investment problems: a synthesis *Paris-Priceton Lectures on Mathematical Finance 2002, Springer Lecture Notes in Mathematics* **1814**, 95–131.

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

SCHMOCK, U., SHREVE, S. E. & WYSTUP, U. (2001), Valuation of exotic options under shortselling constraints, *Finance & Stochastics*, to appear.

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.